DEVELOPMENT AND EVALUATION OF REVISED CLASS C DRIVER LICENSE WRITTEN KNOWLEDGE TESTS

By
Eric A. Chapman & Scott V. Masten

May 2002

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.			
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE May 2002	3. REPORT TYPE AND DATES COVERED Final Report	
4. TITLE AND SUBTITLE Development and Evaluation of Revised Class C Driver License Written Knowledge Tests			5. FUNDING NUMBERS
6. AUTHOR(S) Eric A. Chapman and Scott V. Masten			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) California Department of Motor Vehicles Research and Development Section P.O. Box 932382 Sacramento, CA 94232-3820			8. PERFORMING ORGANIZATION REPORT NUMBER CAL-DMV-RSS-02-196
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILABILITY STATEMENT			12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report represents the results of an evaluation of English and Spanish language Class C license written knowledge examinations administered to applicants for an original or renewal driver license.. The tests were extensively modified following the 1999 statewide evaluation (Masten, 1999). The study assessed the fail rate, mean number of errors, and internal-consistency reliability for each test form, as well as the pass rate, percentage of applicants selecting each answer choice, and item-total correlation for each item on each English language test form. The results are based on 10,502 completed test forms that were collected from field offices statewide in April 2001. It was found that the test fail rates for all tests decreased from the last statewide evaluation. However, the disparity in fail rates between the English and Spanish tests increased, with the rates for Spanish applicants continuing to be substantially higher than the rates for English applicants.

14. SUBJECT TERMS fail rates, reliabilities, item statistics, written test			15. NUMBER OF PAGES 49
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT None

NSN 7540-01-280-5500

Standard Form 298 (2-89)
Prescribed by ANSI Std. Z39-18
298-102

PREFACE

This report is issued as an internal monograph of the Department of Motor Vehicles' Research and Development Branch rather than an official report of the State of California. The findings and opinions may not represent the views and policies of the State of California.

ACKNOWLEDGMENTS

This project was conducted under the general direction of Cliff Helander, Research Chief, and the supervision of Robert Hagge, Research Manager II. Martha Boudreau, Manager III, provided specific information in relation to the written test content and format. Staff in the Driver License Issuance Branch, under the supervision of Carmen Rios, Manager IV, key entered data from the completed written test sheets. Debbie McKenzie, Associate Government Program Analyst, and Carolina Mundarain, Student Assistant, helped type and proofread the report.

SUMMARY

<u>Introduction</u>
- This report presents the results of an evaluation of the English DL 5 (Rev. 2/01), Spanish DL 5 (Rev. 3/01), and English DL 5T (Rev. 2/01) Class C license written knowledge examinations. The study assessed the fail rate, mean number of errors, and internal-consistency reliability for each test form, as well as the pass rate, percentage of applicants selecting each answer choice, and item-total correlation for each item on each test form for the English tests. Also presented is an assessment of the randomness of the answer choice assignment for the English DL 5.

- The tests were extensively modified following the 1999 statewide evaluation (Masten, 1999). One of the objectives of the current evaluation was to assess the effects of these changes, particularly with regard to the test fail rates. Because of the substantial alterations made to the test items, including expanding the item pool, it was anticipated that this analysis would identify many items that need to be modified or replaced.

- The results are based on 10,502 completed test forms that were collected from field offices statewide on April 26, 2001 or, in a few instances, a subsequent Thursday.

Results
- The fail rates for the English language original, renewal, and provisional tests and the Spanish language original and renewal tests are lower than those reported in the 1999 statewide written test evaluation. The overall fail rate for applicants on their first-test attempt is 50.4% for English originals, 31.1% for English renewals, 80.3% for Spanish originals, 71.2% for Spanish renewals, and 48.3% for English provisionals. The differences between the 1999 and 2001 written test evaluation fail rates for these groups are illustrated in Figure 1.

- The fail rates for all applicant groups tend to increase or stay consistent on the second and third test attempts. This finding is discouraging because it suggests that applicants are not reviewing the driver license handbook before retaking the test. The fail rates for the three attempts on each test are shown in Figure 2.
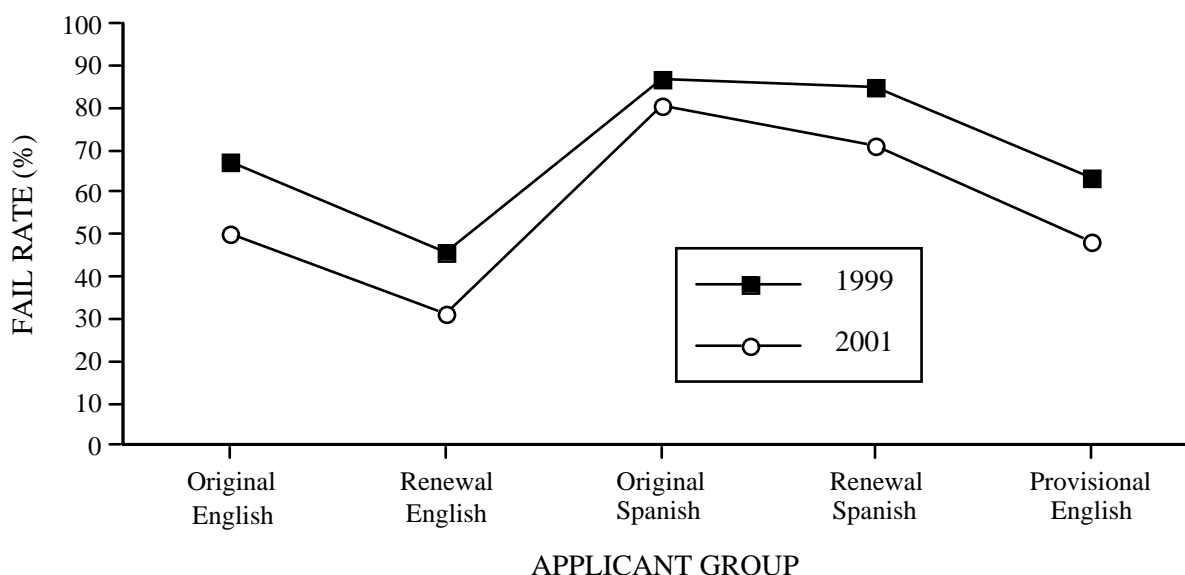

Figure 1. 1999 and 2001 written test evaluation first-attempt fail rates for English and Spanish original and renewal applicants and English provisional applicants.
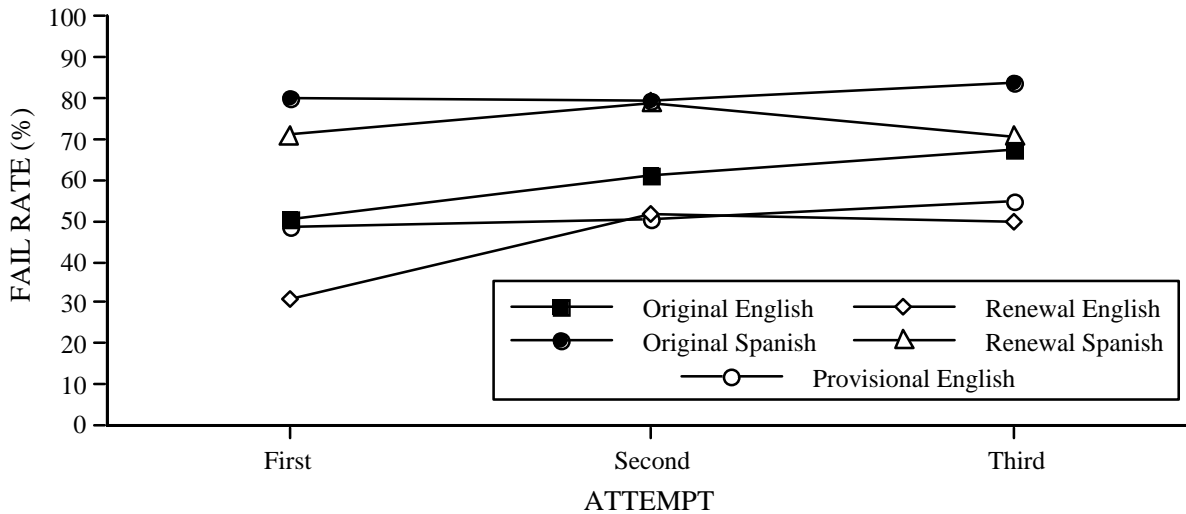
Figure 2. Test fail rates for English and Spanish original and renewal applicants and English provisional applicants by attempt number.

- Some forms of the same test are more difficult than others, differing by as much as 34 percentage points.

- Almost all of the internal-consistency test reliabilities for the English and Spanish DL 5 test forms for original applicants and the English DL 5T test forms for provisional applicants are adequate, and some are even "good" or "excellent." Any test form with a reliability below .70 is inadequate and should be reviewed and revised to improve its reliability. Test reliability can be increased and made more equal across different forms of the same test by correcting or replacing problem items, particularly those with low or negative item-total correlations.

- None of the English DL 5 and only two of the Spanish DL 5 test form reliabilities for renewal applicants exceeded the .70 whole-test reliability standard of acceptability. Overall, the renewal test reliability findings indicate that 18 items are too few to produce renewal driver license knowledge tests with adequate reliability.

- Several of the items on each test form are potentially deficient as indicated by a low item-total correlation, a pass rate that is too high or too low, or a distractor selection

rate that is too high or too low. Items with poor item-total correlations are the most suspect and special care should be taken to review and correct them.

- Answer choices "a," "b," and "c" were equally represented as correct answers on both the original and renewal English DL 5 tests, which is desirable.

- Some field offices administered the back of the DL 5 test sheet to renewal applicants, which is inconsistent with department policy to use only the front of the form to test renewals.

- Examiners often did not count all incorrect items when calculating the total test score for original, renewal, and provisional applicants. This resulted in the computer-graded fail rates reported in this evaluation being slightly higher than the true operational fail rates. The true operational fail rate across all test attempts is 47.0% for English originals, 24.7% for English renewals, and 44.6% for English provisionals.

- Many field offices were using older rather than current revisions of the Class C license tests. This reduces the effectiveness of the current practice of randomizing the order of items on the English DL 5 every 3 months to curtail applicant cheating.

- The readability level of both the English and Spanish versions of the written test were found to fall at or below the sixth-grade reading level. A reading level of fifth- or sixth-grade is considered optimum, by AAMVA standards, in communicating with driver license applicants who have difficulty reading.

Recommendations
- The tests should be reviewed and revised where appropriate to further improve their quality. In particular, the different forms of the same test should be made more equal in difficulty level and reliability, and items with characteristics that indicate they may be deficient should be reviewed and modified or replaced as necessary. The following three recommendations should help accomplish these goals.

  1. Questions with item-total correlations that are below .10 or negative should be reviewed and almost always need to be modified or replaced because they are usually poorly worded and confusing. Item distractors (incorrect answer choices) that none, or almost none, of the applicants chose should also be rewritten or replaced.

2.   Items with pass rates that are too high or too low, or with distractor selection rates that are too high or too low, should be reviewed for possible problems and modified as necessary. The low pass rate items may have confusing wording and, if so, should be rewritten to make them clear.

3.   The order of answer choices within each item on each test form should be randomized periodically to decrease an applicant's chance of guessing the correct answers. There are computer applications available that can efficiently and cheaply accomplish this goal. Currently only the order of items on each form is randomized.

- The reliabilities of the renewal tests are too low and should be increased. The department can accomplish this by increasing the length of the renewal tests by having renewal applicants complete all 36 items on the DL 5 instead of only the first 18. It is estimated that doubling the test length would increase the overall English DL 5 renewal test internal-consistency reliability from .55 to .71, and the overall Spanish DL 5 renewal test reliability from .68 to .81. Correcting problem items, especially those with a low item-total correlation, would also help to increase test reliability, though probably not to the extent necessary without also increasing the number of test questions.

- Steps should be taken to ensure that field office personnel are administering only current versions of the tests in accordance with procedures stated in the Driver License Manual. Doing so would increase the effectiveness of the current practice of randomizing the English DL 5 test every 3 months to reduce the possibility of cheating. This deviation from procedures has been found consistently in prior test evaluations. Clearly the efforts used in the past to correct this practice are inadequate.

- The department should continue investigating the possible use of computer technology for generating the test forms from a large item-pool database and possibly creating a unique test for each applicant.

- The department should collect a larger sample of Spanish language tests during the next statewide evaluation. This would make it possible to calculate item statistics for both language groups.

- The department should consider the following strategies for increasing the level of knowledge of the applicants and thereby improving test scores:

    1. Prepare and distribute information brochures and press releases that publicize the knowledge content areas and principles that applicants most frequently have problems with and emphasize the importance of reading the driver handbook before taking the test.

    2. Develop and implement procedures that would better identify applicants who need to take an oral rather than written knowledge test. The oral test aids low-literacy applicants who have the necessary knowledge of safe driving practices and rules of the road, but have difficulty reading the test.

    3. Require waiting periods between test attempts, which is already being done for provisional license applicants. This would increase the incentive of applicants to read and study the handbook more thoroughly before their first and subsequent test attempts.

    4. The department should stop the policy of returning graded test forms to encourage applicants to read the manual rather than relying only on reviewing the old test forms to prepare for the next test.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

APPENDICES

LIST OF TABLES

TABLE OF CONTENTS (continued)

LIST OF FIGURES

INTRODUCTION

This report presents results of an evaluation of the English DL 5 (Rev. 2/01), Spanish DL 5 (Rev. 3/01), and English DL 5T (Rev. 2/01) Class C license written knowledge examinations. The data for this report were collected as part of the Department's annual statewide evaluation of the written tests. The purpose of this report is to provide test and item statistics to be used for making the test forms more reliable and valid testing instruments.

There are 10 different versions or forms of the English DL 5 examination and five forms of the Spanish DL 5 examination, each consisting of 36 items. These tests are administered to Class C license applicants who are 18 years of age or older. License renewal applicants complete only the first 18 items, while applicants for an original license complete all 36 items. There are five different forms of the English DL 5T examination, each consisting of 46 questions. The DL 5T is administered to original Class C license applicants who are younger than 18 years of age (provisional licensees).

The last statewide written test evaluation was published in March 1999 (Masten, 1999). At about the same time, the American Association of Motor Vehicle Administrators (AAMVA) released their *AAMVA Guidelines for Knowledge and Skill Testing* (1999), which provides guidelines for the development and evaluation of written knowledge and drive tests for license applicants. The AAMVA written knowledge test guidelines include the following recommendations:

- Randomly select the questions for each test form from a large pool of items covering the entire manual to prevent applicants from being able to pass the test merely by memorizing the answers to questions on old test forms.

- Sample content for each test form representatively across all knowledge categories to provide the best estimate of the applicant's total knowledge attainment. In addition, do not put the same question on more than one test form, and try to make all alternate forms equal in difficulty level.

- No question on any one form should provide information that would give the answer to another item on the form.

1

- The content of the knowledge tests should be drawn directly from the driver license handbook, cover all sections of the handbook, and to the extent possible be worded in the same manner as in the handbook.

- Incorrect responses for an item should address the same content as the correct answer and be incontrovertible incorrect.

- There should be only one correct answer to each question and the position of the correct answer among all answer choices should be randomized across test items.

- The questions should be worded to avoid use of complex words or phrases, "all of the above" or "none of the above" answer choices, "legalese," the negative form (e.g., "which of the following is not…"), inconsistent answer choices such as those that are much longer or shorter in length than the others in the item, ambiguous wording, and unnecessarily fine distinctions between answer choices.

- The tests and driver handbook should be written at a fifth- to sixth-grade reading level.

In April 1999, the Department began the process of completely revising the English and Spanish language written tests so that they would be consistent with the AAMVA standards. Knowledge experts in the Department created a pool of 342 new test questions based solely on the content in the *California Driver Handbook*. The new item pool provided enough questions so that each of the 10 forms of the English DL 5 would otherwise have unique items, with the exception of two items (related to blood alcohol content and days to report the sale or transfer of a vehicle) that are required by law to be included on all test forms. Items were created for 23 different content areas. Items from each content area were assigned to the 10 forms of the English DL 5 in equal proportions. The content areas and number of items that were assigned to each of the 10 test forms are shown in Table 1.

Table 1

Knowledge Content Areas on the English DL 5 Written Tests and
Assignment of Items

Knowledge content area	Total number of items	Number of items on each test
Accident responsibility	10	1
Driving under the influence	10	1
Driving on freeways	10	1
Lane markings	10	1
Lane usage	10	1
Mandatory questions (BAC & vehicle sale)	2	2
Parking on hills	10	1
Parking (general)	10	1
Road hazards	10	1
Railroad crossings	10	1
Right-of-way	20	2
Seat belts and child restraints	10	1
Space cushion (around vehicle)	20	2
Speed and speed limits	20	2
Safe driving practices[a]	30	3
Sharing the roadway with others	10	1
Driving with special vehicles	10	1
Improving traffic flow	10	1
Traffic lights and signals	20	2
Turns	20	2
Traffic signs	30	3
Visual scanning	30	3
Driving in inclement weather	20	2
Total	342	36

Note. There are 10 equivalent forms of the English-language written test. Although not an official content area, each test form contains two or more items that relate to the interaction of vehicles and pedestrians.

[a]This category contains items relating to vehicle equipment usage (e.g., horn, headlights, turn signals, parking lights, and emergency flashers), general safe driving rules, accident avoidance and protection, defensive driving, driving when tired, and other subject matter.

The 10 new test forms were pilot tested in October 1999 in the 20 DMV field offices with the highest license transaction volumes. Each office was given 200 tests (20 of each form) to administer to first-attempt English-test applicants instead of the then current revision of the DL 5 test. The offices returned the completed tests to DMV headquarters. The test items were analyzed and then, when necessary, revised, replaced, or otherwise modified based on the statistical results calculated for each item. Replaced items were always substituted by another item from the same content area to maintain the original content area assignment ratios. The resulting revised tests were again pilot tested in April 2000 in the same 20 offices using the same procedures. Focus groups were also conducted to gather opinions about the test content and readability. Item and test statistics from the April 2000 pilot, along with information gathered from the focus groups, were used to further revise the items where necessary to improve the overall reliability and validity of the tests. A final revised test was released statewide in February 2001. This version (Rev. 2/01) of the English DL 5 is evaluated in this report. The first 5 forms of the English DL 5 were translated into Spanish by department translators and released statewide in March 2001 as the Spanish DL 5 (Rev. 3/01). In addition, the items from the English DL 5 were used to make new versions of the DL 5T (Rev. 2/01) for provisional license applicants, which were released statewide in February 2001.

A change in scoring procedures was also enacted with the release of the 2001 tests to equalize the expected knowledge competency of original, renewal, and provisional applicants. Specifically, the number of errors allowed was increased from 5 to 6 (out of 36 total) for adult original applicants and from 7 to 8 (out of 46 total) for provisional license applicants. This change resulted in an 83% expected competency level (i.e., the percentage of items that must be answered correctly) for adult original and provisional tests, which is the same as the existing level for the renewal tests.

This report presents results of the first statewide evaluation of these revised tests. The objective of this evaluation was to assess the effects on test fail rates of changing the number of errors allowed on the original and provisional tests, and to provide test and item statistics to guide the DMV in further improving the reliability and validity of the tests.

This report presents the fail rate, mean number of errors, and internal-consistency reliability for each form of the English and Spanish DL 5 tests and the English DL 5T

test. Results for the English and Spanish DL 5 tests are presented separately for original and renewal applicants, and all test fail rates are presented separately for first, second, third, and fourth or higher attempts on the tests. In addition, the pass rate, percentage of applicants choosing each answer choice, and item-total correlation for each item on each test form, and also the percentage of applicants who would pass each test form at different cut-points, are presented for the English DL 5 and English DL 5T tests. All item statistics are based on first-attempt applicants (those taking the test for the first time on the current application). Item statistics are not presented for the Spanish DL 5 tests because too few of the tests were collected to produce accurate estimates. Neither item nor test statistics are presented for the Spanish DL 5T (Rev. 6/00), again because an insufficient number of these tests were collected.

<div align="center">METHODS</div>

Data Collection

All DMV field offices were instructed to send to headquarters all Class C driver license written knowledge tests completed on Thursday, April 26th, 2001. Some offices failed to submit data for that day and were asked to make up for it by collecting all tests completed on a subsequent Thursday. Tests were ultimately received from 167 of the 168 field offices that were open during the data collection period. The one office that did not submit any data indicated that they had no license applicants on the day of data collection.

The field offices were instructed to administer and collect only the most current revisions of the tests, score the tests in the usual manner, circle the correct answers to missed questions, and write the total score (the number of wrong answers) and field office number on the front of the test sheet. The offices were also instructed to write "1st," "2nd," "3rd," "4th," and so forth on the bottom of the front of each test form to indicate the test attempt on the current license application, and to write "ORIG" on the bottom of the front of each test if the applicant was applying for an original license or "REN" if they were applying for a renewal license. During the data collection period, tests were not returned to applicants. However, if an applicant requested their test back, field office personnel were instructed to give them a photocopy of the test.

The tests were screened by the Research and Development Branch (R&D) and key entered into an electronic file by the Driver License Issuance Branch. The tests were graded by computer to ensure that test and item fail rates were computed correctly.

Data Analysis

A statistical technique known as analysis of variance (ANOVA) was used to determine whether any differences between the test form fail rates or average scores were statistically significant. Differences are considered significant if the probability (p) of their occurrence by chance alone is less than 5 times in 100. Games and Howell multiple-comparison tests were used for post hoc comparison tests when a significant omnibus ANOVA was found. These tests determined which specific test form pairs were significantly different from one another on test fail rate.

The internal-consistency reliability of each test form was computed using the Kuder-Richardson formula (K-R 20). In general, this type of reliability indicates the degree of uniformity among test items and the extent to which the test items measure a similar domain of knowledge. It also serves as a gauge of the overall precision of the test as a measurement instrument. A test that is highly reliable is likely to result in very similar scores across repeated testing of the same people (assuming that their true knowledge level doesn't change between test administrations). The reliability coefficient can range from 0 to 1, where a value of 0 indicates no similarity between the test items and a value of 1 denotes that the items are perfectly homogeneous in content. Coefficients closer to 1 are more desirable, and any test with a coefficient below the .70 whole-test standard is probably in need of revision.

The item-total correlation coefficient measures the degree to which performance on an individual item is related to performance on the entire test. The coefficient can theoretically range from -1 to 1, with the zero midpoint value representing no relationship at all between performance on the item and performance on the test. In other words, performance on an item with a coefficient approaching zero (e.g. between -.10 and .10) has very little or no relationship to whether applicants scored high or low on the overall test. Items with a positive coefficient value are more likely to have been answered correctly by applicants who scored higher on the test, while items with a negative coefficient value are more likely to have been answered correctly by applicants who scored lower on the test. As such, negative item-total correlations are almost always undesirable.

RESULTS

Data Collection and Screening

A total of 13,235 test forms of various types were received by R&D for the 1-day collection period. Of these, a total of 2,733 tests were removed from the evaluation because they were in a language other than English or Spanish (1,060), not an English or Spanish DL 5 or DL 5T (813), not the most recent revision of the test (606), had a form number that could not be accurately determined (204), or because renewal applicants were administered the back of the DL 5 test instead of the front, which violates departmental policy stated in the Driver License Manual (50). The screening process resulted in 10,502 usable test forms for the evaluation.

Ordinarily it is desirable that at least 100 first-attempt test sheets of a given test form be analyzed to produce reasonably accurate estimates of item statistics for that form. This standard was met for the English DL 5 (Rev. 2/01) for original and renewal applicants, and also for the English DL 5T (Rev. 2/01) for provisional applicants. However, this standard was not met for the Spanish DL 5 (Rev. 3/01) for originals and renewals, so item statistics were not calculated for these tests.

All usable test forms were graded by computer, with an item being counted as incorrect if the wrong answer choice was marked, the item was left blank, or more than one answer choice was marked. The computer-graded test scores were used to compute the test fail rates and internal-consistency reliabilities. All fail rates presented within this report are based on the current minimum passing standards allowing six errors (out of 36 test items) for DL 5 original applicants, three errors (out of 18 test items) for DL 5 renewal applicants, and eight errors (out of 46 test items) for DL 5T provisional applicants.

Examiner Scoring Bias

Computer grading of the tests revealed that the actual number of errors made by the applicant often differs from the error score written by the examiner on the front of the test form. The scoring bias for these forms is almost always in the applicant's favor, and often affects the test result. To demonstrate the difference between the computer-graded and examiner-graded fail rates, three different fail rates are presented in Table 2 for English DL 5 original and renewal applicants and English DL 5T provisional

applicants. The first fail rate was calculated from computer grading of the tests. The second fail rate was calculated from the scores that the field office examiners wrote on the tests, which were not available for all the tests. The third fail rate represents a blending of the previous two rates; the examiner score was used to calculate the fail rate if a score was available, otherwise the computer-graded score was used. The differences in the fail rates are illustrated in Figure 3.

Table 2

Number of Tests (*n*) and Fail Rate for First-Attempt English Original, Renewal, and Provisional Applicants when Tests were Graded by Computer, Examiner, and Examiner/Computer in Combination

Test type	Computer graded		Examiner graded		Examiner/computer graded	
	n	Fail rate (%)	*n*	Fail rate (%)	*n*	Fail rate (%)
English DL 5 original	2,005	50.4	1,812	47.0	2,005	46.6
English DL 5 renewal	2,831	31.1	2,310	24.7	2,831	26.3
English DL 5T provisional	911	48.3	830	44.6	911	45.0

Note. Examiner/computer grading used the examiner score if available and the computer score otherwise.

The difference between the examiner and computer scores is primarily due to some examiners' having discussed missed items with examinees and having awarded points based on the latter's verbal responses. At the time test forms were being collected for this evaluation, department policy allowed this practice for all applicants except provisionals. (The differences between the computer and examiner fail rates for provisional applicants indicates that some field office personnel were also inappropriately applying this practice to minors.) This examiner scoring bias in the applicant's favor has caused the test fail rates presented in this report, which are based on computer-graded scoring, to be slightly higher than the true operational fail rates occurring in the field offices. The true operational (examiner-graded) fail rate across all test attempts is 47.0% for English originals, 24.7% for English renewals, and 44.6% for English provisionals. It should be noted that the policy of allowing examiners to rephrase or restate missed questions was repealed on June 18, 2001, which would be expected to lead to a slight increase in the operational test fail rates statewide.
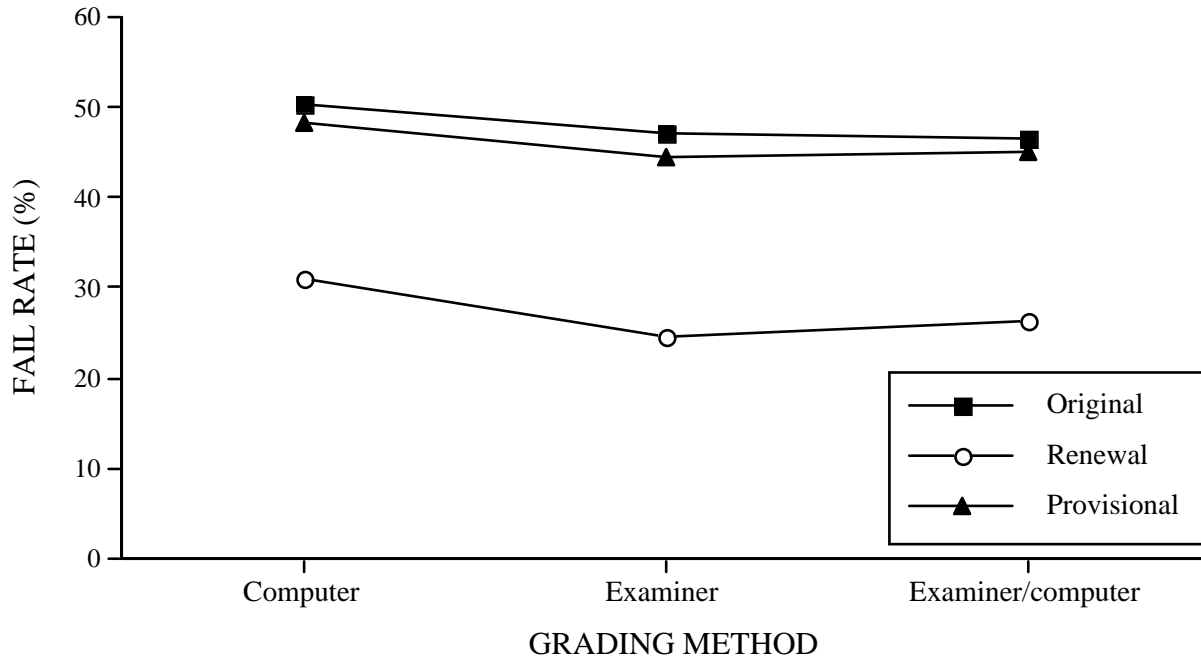
Figure 3. Fail rates for English original, renewal, and provisional first-attempt applicants by method of grading.

Test Statistics

Test Form Difficulty and Reliability

The number of tests given, fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form and applicant type are presented in Table 3. The differences in the form fail rates and mean errors, and the pattern of internal-consistency reliabilities for the forms, are illustrated for each applicant type in Figures 4, 5, and 6, respectively. The test statistics shown in the table and figures are discussed in the following five subsections.

English DL 5 original applicants. For original applicants taking the English DL 5, the average test form fail rate was 50.4% and the 10 form fail rates range from 42.4% to 59.3%. The differences between the form fail rates are statistically significant ($p < .01$), with Form 10 having a higher rate than Forms 1 and 9, and Form 5 having a higher rate than Form 9. The differences in fail rates for the other possible form pairs are not significant.

9

Table 3

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for each Form of the English DL 5 (Rev. 2/01), Spanish DL 5 (Rev. 3/01), and English DL 5T (Rev. 2/01) for First-Attempt Applicants

Test form	*n*	Fail rate	Mean errors	Reliability
Original DL 5 English[a]				
1	210	43.8	6.6	.77
2	190	51.6	7.4	.76
3	206	46.6	6.8	.72
4	214	44.9	7.0	.78
5	207	58.9	8.1	.78
6	201	52.2	7.8	.82
7	188	46.8	6.9	.76
8	208	55.8	7.8	.77
9	172	42.4	6.9	.82
10	209	59.3	8.1	.76
Total	2,005	50.4	7.4	.77
Renewal DL 5 English[b]				
1	305	30.2	2.9	.49
2	314	39.8	3.2	.56
3	264	26.1	2.4	.48
4	262	22.9	2.4	.62
5	274	46.7	3.4	.48
6	259	24.3	2.5	.53
7	281	16.4	2.1	.57
8	288	50.3	4.0	.60
9	306	21.9	2.4	.61
10	278	30.9	2.9	.55
Total	2,831	31.1	2.8	.55
Original DL 5 Spanish[c]				
1	81	77.8	11.2	.82
2	80	77.5	10.7	.72
3	75	78.7	12.7	.86
4	73	91.8	14.3	.77
5	86	76.7	11.9	.86
Total	395	80.3	12.1	.81
Renewal DL 5 Spanish[d]				
1	29	55.2	4.8	.61
2	33	75.8	6.0	.69
3	28	71.4	6.0	.83
4	20	75.0	6.0	.78
5	22	81.8	6.0	.48
Total	132	71.2	5.7	.68
Provisional DL 5T English[e]				
1	200	46.5	8.5	.77
2	189	32.3	7.7	.74
3	166	54.8	10.2	.83
4	176	54.6	9.6	.69
5	180	55.0	9.7	.78
Total	911	48.3	9.1	.76
Grand total	6,274	43.8	5.8	.67

Note. The figures presented for total and grand total fail rate, mean errors, and reliability are weighted averages. All ANOVAs are two-tailed.
[a]Forms differ significantly on fail rate ($F = 3.17$, $p < .01$) and mean errors ($F = 3.08$, $p < .01$).   [b] Forms differ significantly on fail rate ($F = 17.48$, $p < .001$) and mean errors ($F = 20.77$, $p < .001$).   [c]Forms differ significantly on fail rate ($F = 4.69$, $p < .01$) and mean errors ($F = 5.30$, $p = .001$).   [d]Forms did not differ significantly on fail rate ($F = 1.33$, $p = .26$) or mean errors ($F = 0.77$, $p = .55$). [e]Forms differ significantly on fail rate ($F = 7.32$, $p < .001$) and mean errors ($F = 7.16$, $p < .001$).
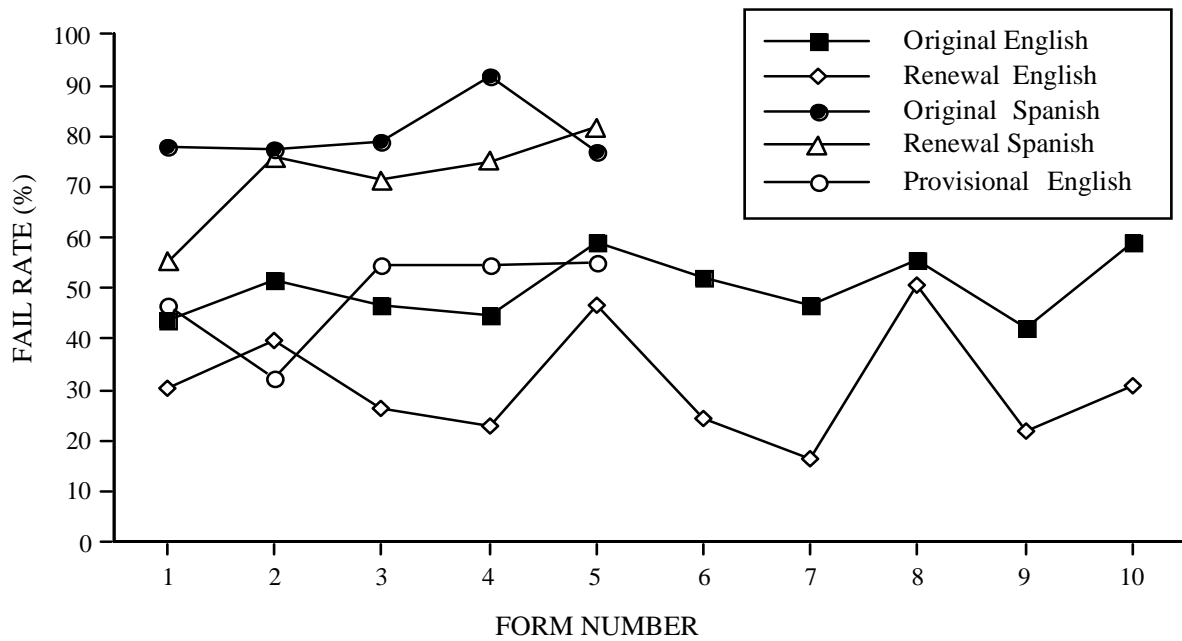
Figure 4. Form fail rates for first-attempt English and Spanish DL 5 original and renewal applicants and English DL 5T provisional applicants.
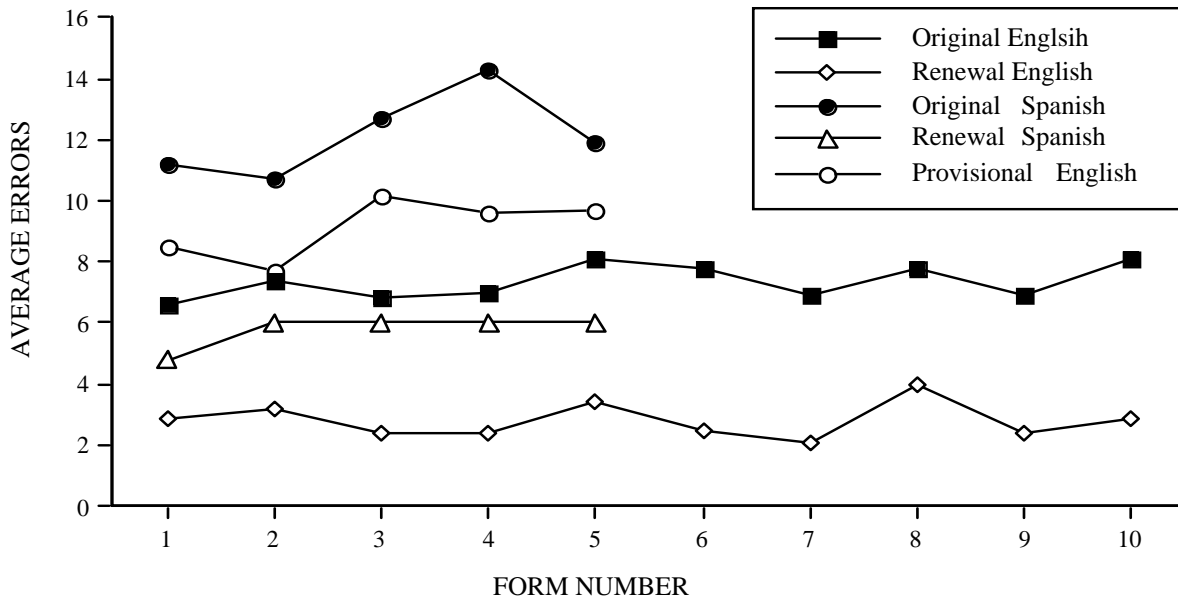
Figure 5. Form mean numbers of errors for first-attempt English and Spanish DL 5 original and renewal applicants and English DL 5T provisional applicants.
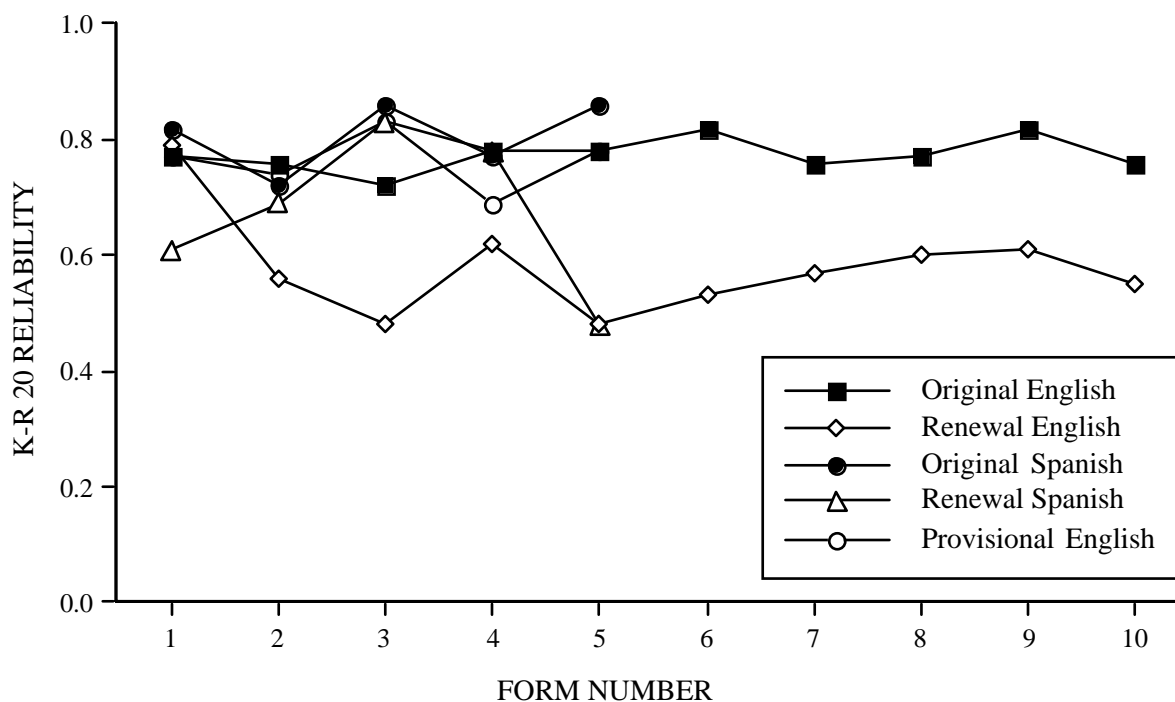
Figure 6. Form internal-consistency reliabilities for first-attempt English and Spanish DL 5 original and renewal applicants and English DL 5T provisional applicants.

The mean number of errors was 7.4 overall. The values for the 10 forms range from 6.6 to 8.1, and the differences are statistically significant ($p < .01$). Applicants who completed Form 5 or 10 made significantly more errors than did applicants who completed Form 1. Differences between the mean errors for the other form pairs are not significant.

The K-R 20 reliabilities for the 10 forms range from .72 to .82. All of these coefficients exceed the .70 whole-test reliability standard.

English DL 5 renewal applicants. The fail rates for the 10 test forms of the English DL 5 for renewal applicants range from 16.4% to 50.4% and average 31.1% overall. The differences between the form fail rates are statistically significant ($p < .001$). The fail rates for Forms 5 and 8 are significantly higher than the rates for all the other forms except Form 2. The fail rate for Form 2 is significantly higher than the rates for Forms 3, 4, 6, 7, and 9. In addition to the differences already mentioned, the fail rates for Forms 1 and 10 are significantly higher than the rate for Form 7. The differences between the other form fail rates are not significant.

The mean number of errors was 2.8 across the 10 forms. The mean scores on the individual forms range from 2.1 to 4.0, and the differences are significant ($p < .001$). Applicants who completed Form 8 missed significantly more items than did applicants who completed any of the other test forms except Form 5. Applicants who completed Form 5 missed significantly more items than did applicants who completed Forms 1, 3, 4, 6, 7, or 9; those who completed Form 2 missed significantly more items than did those who completed Forms 3, 4, 6, 7, or 9; and those who completed Forms 1 or 10 missed significantly more items than did those who completed Forms 7. Differences between the other form mean errors are not significant.

The K-R 20 reliabilities for the 10 forms range from .48 to .62, all falling below the .70 whole-test standard. The fact that these reliability coefficients are much lower than those for the 36-item test is due primarily to the difference in test length rather than to any defect in the content of the items. The results indicate that 18 items may be too few to produce an adequate level of test reliability for a driver license knowledge test.

Spanish DL 5 original applicants. The fail rates for the five forms of the Spanish DL 5 for original applicants average 80.3% and range from 76.7% to 91.8%. The differences between the fail rates are not significant ($p > .10$).

The mean number of errors for the forms average 12.1 and range from 10.7 to 14.3, and the differences between forms are significant ($p < .01$). Applicants who completed Form 4 missed significantly more items than did applicants who completed Forms 1 or 2. Differences between the mean errors for the other forms are not significant.

The five form K-R 20 reliabilities range from .72 to .86. All these reliabilities fall above the .70 whole-test standard.

Spanish DL 5 renewal applicants. The fail rates for the five forms of the Spanish DL 5 for renewal applicants average 71.2% and range from 55.2% to 81.8%. The differences between the form fail rates are not significant ($p > .25$).

The mean number of errors for the forms average 5.7 and range from 4.8 to 6.0, and the differences between forms are also not significant ($p > .50$).

The five form K-R 20 reliabilities range from .48 to .83. Three of the five form reliabilities fall below the .70 whole-test standard. These results once again indicate that

18 items may be too few to produce driver license knowledge tests with adequate reliability.

English DL 5T provisional applicants. The fail rates for provisional applicants taking the English DL 5T average 43.8% and range from 32.2% to 55.0%. The differences between the form fail rates are significant ($p < .001$). The fail rate for Form 2 is significantly lower than the rates for all other forms. The differences between the other form fail rates are not significant.

The mean number of errors for the five test forms of the English DL 5T for provisional applicants average 9.1 and range from 7.7 to 10.2, and the differences are significant ($p < .001$). Applicants who completed Form 2 made significantly fewer errors than did applicants who completed Forms 3, 4, and 5, and those who completed Form 1 missed fewer errors than did applicants who completed Form 3. The differences between the other form mean errors are not significant.

The five form K-R 20 reliabilities range from .69 to .83. Only one form has a reliability that falls below the .70 standard.

Test Difficulty by Attempt Number
The number of tests administered (n), fail rate, and mean number of errors for applicants attempting a test for the first, second, third, or fourth or more time or who did not have an attempt number recorded on the form are presented in Table 4. The non-reporting of attempt number is not likely to be associated with test performance, and therefore the exclusion of cases with a missing attempt number is not believed to have significantly biased the attempt fail rate estimates.

The test fail rate averaged over all attempts is 55.8% for English DL 5 original applicants, 35.1% for English DL 5 renewal applicants, 80.7% for Spanish DL 5 original applicants, 73.5% for Spanish DL 5 renewal applicants, and 49.4% for English DL 5T provisional applicants. The fail rates generally increased or stayed the same over successive attempts, indicating that many of these applicants either do not prepare for subsequent test attempts by reading the driver handbook or have difficulty reading the handbook or tests.

Table 4

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for the English DL 5 (Rev. 2/01),
Spanish DL 5 (Rev. 3/01), and English DL 5T (Rev. 2/01) by Test Attempt Number

Test and Attempt	*n*	Fail rate (%)	Mean errors
Original DL 5 English			
First	2,005	50.4	7.4
Second	877	61.0	8.4
Third	405	67.2	8.7
Fourth or higher	104	62.5	8.7
Missing	475	58.3	8.1
Total	3,866	55.8	7.9
Renewal DL 5 English			
First	2,831	31.1	2.8
Second	421	51.5	3.9
Third	140	50.0	4.2
Fourth or higher	17	64.7	5.5
Missing	418	39.5	3.7
Total	3,827	35.1	3.1
Original DL 5 Spanish			
First	395	80.3	12.1
Second	303	79.5	11.1
Third	146	83.6	10.9
Fourth or higher	41	73.2	10.6
Missing	88	85.2	12.0
Total	973	80.7	11.6
Renewal DL 5 Spanish			
First	132	71.2	5.7
Second	76	78.9	6.0
Third	24	70.8	5.1
Fourth or higher	11	81.8	6.9
Missing	25	68.0	5.2
Total	268	73.5	5.7
Provisional DL 5T English			
First	911	48.3	9.1
Second	330	50.6	9.2
Third	131	55.0	9.6
Fourth or higher	23	47.8	9.5
Missing	173	48.6	9.1
Total	1,568	49.4	9.2
Grand total	10,502	50.2	6.6

Note. All figures presented for total fail rate and total mean errors are weighted averages. Missing-attempt cases
did not have the applicant's attempt number written on the front of the test form. Estimates based on a sample size
of fewer than 20 cases are likely to be unreliable and therefore should be interpreted with caution.

Test Fail Rates by Field Office

The number of tests received and the test fail rates for English DL 5 original and renewal applicants over all test attempts in each field office are presented in Appendix A. Field office fail rates that are computed from fewer than 20 test forms are probably unreliable, and therefore may not be trustworthy. Field office fail rates are not presented for Spanish DL 5 original and renewal applicants, or for English DL 5T provisional applicants, because too few test forms were collected to compute accurate estimates for the majority of the field offices. The English test fail rates for offices with 20 or more forms range from 20.8% to 85.7% for original applicants, and 11.1% to 71.4% for renewal applicants. These numbers indicate wide variation in the knowledge level of applicants residing in different geographical areas of the state.

Test of Answer Choice Randomness

The number and percentage of items for which choice "a," "b," or "c" is the correct answer on each form of the English DL 5 are presented in Table 5 for original and renewal applicants. If the assignment of the correct answers was truly randomized across the answer choices, it would be expected that choices "a," "b," and "c" would each be the correct answer for about 33% of the items across the entire item pool and also on each test form.

For original applicants taking the English DL 5, the percentages of items for which choices "a," "b," and "c" are the correct answers did not differ significantly for all forms combined ($p = .46$), and the differences between the percentages for the 10 test forms were not statistically significant ($p = .44$).

For renewal applicants taking the English DL 5, the percentages of items for which choices "a," "b," and "c" are the correct answers did not differ significantly for all forms combined ($p = .27$), and the differences in the percentages for the three answer choices did not differ significantly between the 10 test forms ($p = .58$).

Table 5

Number (*n*) and Percentage of Times that Each Answer Choice
was the Correct Answer on Each Form of the English DL 5 (Rev.
2/01) for Original and Renewal Applicants

| Test form | Number of test items | Answer choice | | | | | |
| | | a | | b | | c | |
		n	%	*n*	%	*n*	%
Originals[a]							
1	36	14	38.9	11	30.6	11	30.6
2	36	11	30.6	13	36.1	12	33.3
3	36	7	19.4	15	41.7	14	38.9
4	36	12	33.3	15	41.7	9	25.0
5	36	8	22.2	11	30.6	17	47.2
6	36	10	27.8	9	25.0	17	47.2
7	36	10	27.8	14	38.9	12	33.3
8	36	15	41.7	13	36.1	8	22.2
9	36	13	36.1	16	44.4	7	19.4
10	36	13	36.1	14	38.9	9	25.0
Total	360	113	31.4	131	36.4	116	32.2
Renewals[b]							
1	18	6	33.3	7	38.9	5	27.8
2	18	8	44.4	6	33.3	4	22.2
3	18	5	27.8	7	38.9	6	33.3
4	18	6	33.3	8	44.4	4	22.2
5	18	6	33.3	4	22.2	8	44.4
6	18	6	33.3	3	16.7	9	50.0
7	18	5	27.8	7	38.9	6	33.3
8	18	8	44.4	8	44.4	2	11.1
9	18	9	50.0	7	38.9	2	11.1
10	18	8	44.4	6	33.3	4	22.2
Total	180	67	37.2	63	35.0	50	27.8

[a]The percentages of items for which choices "a," "b," and "c" were the correct answers did not significantly differ (χ^2[2, *N* = 360] = 1.55, *p* = .46), and the pattern of these percentages did not significantly vary across the 10 forms (χ^2[18, *N* = 360] = 18.28, *p* = .44). [b]The percentages of items for which choices "a," "b," and "c" were the correct answer did not significantly differ (χ^2[2, *N* = 180] = 2.63, *p* = .27), and the pattern of these percentages did not significantly vary across the 10 forms (χ^2[18, *N* = 180] = 16.13, *p* = .58).

Item Statistics

All item statistics presented here are based on first-attempt applicants to eliminate potential learning effects of previous exposure to the test items. The results of the analyses of test items are presented in Appendices B, C, and D. Appendix B contains the results for the English DL 5 for original applicants, Appendix C contains the results for the English DL 5 for renewal applicants, and Appendix D contains the results for the English DL 5T for provisional applicants. As indicated earlier in this report, item statistics are not presented for the Spanish DL 5 applicant groups because too few tests were received for these applicants to accurately compute the statistics. Appendices A, B, and C each contain four tables that are always presented in the same order. The four tables contain, respectively, item pass rates and answer choice selection rates; item-total correlations; percentage of applicants who would pass at different cut-points; and summary of problem items on each test form. The results presented in these tables in each appendix are described in the next four subsections of this report.

Item Pass Rate and Answer Choice Selection Rates
The percentage of applicants selecting each item answer choice on each test form appears in the first table of Appendices B, C, and D (Tables B1, C1, and D1). The percentages for the correct answer choices are underlined in the tables. The percentages for each item are based on applicants who selected a valid item choice, and excludes applicants who did not choose an answer or chose more than one answer per item. The excluded cases make up less than 3% of applicants taking the test.

The item pass rate refers to the percentage of applicants who correctly answered the item. A very low item pass rate may indicate that the item is poorly worded, has ambiguous or misleading answer choices, is not related to the general knowledge domain that is being tested, or is problematic for some other reason. Items that 60% or fewer of the applicants answered correctly are generally considered suspect, and should therefore be reviewed for clarity and accuracy. Items with extremely high pass rates (95% or higher) are also questionable and should be reviewed as well, because they may not discriminate between people with different levels of knowledge. These "freebie" items often occur when the distractor choices are so illogical that the correct answer is obvious, or when the knowledge required for a question has become commonly known. It is also desirable that the alternative choices be tenable enough to actually attract the

responses of a small but nontrivial percentage of the applicants (generally those who try to guess the answer). Therefore, items with individual distractors that were selected by a very small percentage of applicants, say 2% or less, may also need to be revised. Items with pass rates that are either too high (more than 95%) or too low (less than 60%) are shaded in the tables. Distractors that were selected by 2% or fewer applicants, or that were selected too often (i.e., selected more often than the correct answer or within 10% of the selection rate for the correct answer), are in boldface in the tables.

The 60% and 95% pass rate criteria for items, and the greater than 2% selection rate criterion for distractors, are provided only as guideline indicators of potential item deficiency. Standard item difficulty levels for personnel selection tests generally range from 40% to 70%, while makers of academic tests often attempt to obtain items that 50% of applicants pass. In personnel and academic testing, the purpose is to use the test as a screening device for predicting future achievement, whereas the written driver license knowledge tests are used to encourage applicants to master the information contained in the driver handbook. Hence, the 60% to 95% item difficulty standards used in this evaluation are more relaxed than those used for these other purposes. Although these statistical standards are useful for pinpointing items that may be inadequate, it is not recommended that items be revised or replaced on the basis of these criteria alone. The relative importance of knowledge covered by an item, possible wording problems, and other relevant factors should always be weighed when deciding to revise or replace an item. For example, almost all applicants may have knowledge of certain laws and principles covered by a test question. However, if such knowledge is considered critical to safe driving, the item need not necessarily be discarded simply because 95% or more of the population answered the item correctly.

Item-Total Correlation
The item-total correlation coefficients for the test items are presented in Tables B2, C2, and D2. Items that tended to be answered correctly by applicants who scored low on the test overall (i.e., items with a negative coefficient value), or that had very little or no relationship to the other items on the test (i.e., those with a coefficient value between -.10 and .10), are highly suspect and generally need to be modified or replaced. All items with either of these problems are shaded in the tables.

Percentage of Applicants Who Would Pass at Different Cut-Points

The percentage of applicants who would pass the tests at different cut-points on the tests is presented in Tables B3, C3, and D3. The tables present the percentage of applicants who missed the number of items indicated in the leftmost column of each table row <u>or fewer</u>, and would therefore pass if that number was used as the minimum passing score or cut-point. For instance, Table B3 indicates that 47.1% of original applicants who took Form 1 of the English DL 5 (Rev. 2/01) missed five or fewer items and would have passed if the number of allowable errors had been five, while 56.2% of original applicants taking the same test passed with six or fewer allowable errors. The shaded row in each table shows the pass rate for each form of each test at the current cut-off score of allowable misses (six for DL 5 original applicants, three for DL 5 renewal applicants, and eight for DL 5T provisional applicants).

Summary of Problem Items on Each Test Form

Tables B4, C4, and D4 identify each item with an item-total correlation coefficient value that is less than .10 or negative, a pass rate that is more than 95% or less than 60%, a distractor selected more, or almost, as often as the correct answer or by 2% or fewer of the applicants. Each form has several items with one or more of these characteristics.

DISCUSSION

The test fail rates for all first-attempt applicant groups declined between the 1999 and 2001 statewide evaluations. The rate decline was 17 percentage points for English originals and 15 percentage points each for English renewals and provisionals. For Spanish test-takers, test fail rate decreased by 6 percentage points for originals and 14 percentage points for renewals.

The decreases in test fail rates may be due in part to improvements in the quality of test questions, actions taken by the department to inform applicants of the need to study the driver handbook to pass the test, and allowing one more test questions to be missed on the original and provisional tests.

The frequency distributions showing the number and percentage of applicants missing a specified number of errors or fewer reveals that changing the passing standard for the

original tests (from five to six allowable errors) accounted for 10 percentage points of the 17-percentage point decrease in the fail rate for English originals, and 4 percentage points of the 6-percentage point decline for Spanish originals. The change in passing cutpoint from seven to eight allowable errors for provisionals accounted for 8 percentage points of the 15-percentage point decrease for English provisionals.

It should be noted that the change in the passing standards would be expected to have affected the fail rates for English and Spanish language applicants differently. The reason is that Spanish test-takers miss a much higher number of questions on average than do English test-takers, and therefore the percentage of borderline failures—those failing by only one question—is lower for Spanish than for English applicants. Increasing the number of allowable errors would therefore tend to result in a proportionally higher increase in the pass rate (and greater reduction in fail rate) for applicants taking the test in English than in Spanish.

The passing standard was not changed for renewals and therefore this factor would not help explain the decline in the fail rates for those applicants. However, the fact that English renewals missed fewer questions on average (and had a higher proportion of borderline failures) than did Spanish renewals would have resulted in the former group benefiting more by improvements in the test questions made after the 1999 evaluation.

The difference in performance levels between Spanish and English test-takers is consistent with findings in prior departmental written test evaluations. Several possible factors may help explain the higher fail rate for Spanish applicants, but research has not been conducted to identify these factors and determine the explanatory value, if any, of each one.

While the first-attempt test fail rates may appear discouraging to some, they should not be interpreted to mean that most of those who fail never receive a license. A follow-up review of the driver records of applicants in this evaluation found that the vast majority of English and Spanish language applicants had passed the written test within 9 months following their first test attempt. This suggests that the written test requirement is not a long-term obstacle to obtaining an original or renewal license.

The finding of significant differences in difficulty levels between the test forms is cause for concern because of real and perceived unfairness to individuals who are given the more difficult test forms. Ideally, the forms should be equal in difficulty level, though

this can be hard to achieve. The difference between the highest and lowest form fail rates was 17 percentage points for English originals, 34 percentage points for English renewals, and 23 percentage points for English provisionals. While these differences appear large, they are lower than those found in the 1999 evaluation. Nevertheless, the department should continue to try to make the form fails rates more equivalent.

The English original and provisional tests and the Spanish original tests each have an acceptable level of reliability (.70 or higher). The reliability of the Spanish renewal test across all forms (.68) falls just below the .70 cutoff of acceptability, while the reliability of the English renewal test across all forms (.55) is far below this standard. The lower reliability of the renewal tests can be explained by the fact that they have only 18 items (compared to 36 for originals and 46 for provisionals). The department could increase the reliability of the renewal tests by adding more items. This can be easily achieved by having renewal applicants complete all 36 items on the DL 5 instead of only the first 18 items. Increasing the reliability of the renewal test is important because an unreliable test increases the likelihood that an applicant could get very different test scores when taking the test on two different occasions and consequently reduces the department's ability to make valid decisions on who should and who should not be licensed.

The analysis of individual test questions found that all test forms contained items with an item-total correlation that was too low, a pass rate that is too high or too low, or a distractor that was selected too often or too infrequently. The statistical characteristics of these items indicate that they may be problematic and therefore should be reviewed and revised or replaced as necessary. In particular, items with weak item-total correlations are the most suspect and warrant immediate attention. Items with two or more problem characteristics, for example those with a poor item-total correlation along with a pass rate that is too high or too low, are highly likely to be deficient and should be given special attention in reviewing and correcting them.

The finding that several items have one or more problem indicators is not surprising considering that the 2001 tests have a large number of new items that have never been evaluated psychometrically. However, this is still an improvement over what was found in the 1999 evaluation; an overall comparison of the 1999 and 2001 tests indicate that the latter versions have fewer items with undesirable statistical characteristics.

The item analysis also found that the correct answer was fairly randomly assigned to choices "a," "b," and "c" across the entire item pool, as well as on each of the original

and renewal forms of the English test. This characteristic of the tests will help reduce the chance of an applicant passing the test by trying to guess the correct answers.

During the screening of test forms for this evaluation, it became apparent that some field office personnel were administering the back rather than front side of the DL 5 test sheet to renewal applicants. This deviation from department policy results in applicants not being exposed to the mandatory items pertaining to blood alcohol level and time to report the sale of a vehicle, which appear on the front of each test form. The 50 tests found with this problem were excluded from the analysis of the renewal tests to prevent biasing the results. This is a consistent finding across all of the annual test evaluations.

Some field offices also failed to follow policy by using older revisions of the Class C license test. This practice diminishes the effectiveness of randomizing the English DL 5 tests every 3 months to curtail cheating, and also results in applicants taking an inferior test when improved revisions are available.

Overall, the new tests are much better than the tests evaluated in the 1999 study. There are fewer problem items, the test reliabilities are higher, and the content meets uniform guidelines set forth by AAMVA. These and other factors point to the conclusion that these new tests are more balanced and reliable testing instruments than any prior tests used by the department.

RECOMMENDATIONS

- The department's written tests should be reviewed and revised where appropriate to further improve their quality. In particular, the difficulty level and reliability should be made more equivalent across the different forms of the same test, and items with characteristics that indicate they may be deficient should be reviewed and modified or replaced as necessary. The following three recommendations should help improve the quality of the tests.

  1. Questions with item-total correlation coefficient values that are below .10 or negative should be reviewed and modified or replaced if necessary. Correcting any deficient items is likely to increase the internal-consistency reliability of the tests overall and equalize the reliabilities of different forms of the same test. Item

distractors that none, or almost none, of the applicants chose should also be rewritten to increase test reliability.

2. Items with pass rates that are too high or too low, or with distractor selection rates that are too high or too low, should be reviewed for possible problems and modified as necessary. The low pass rate items should be revised to eradicate confusing wording, and those that have little relevance to safety or are conceptually difficult to grasp should be replaced.

3. The order of answer choices within each item on each test form should be randomized periodically to decrease the applicants' chances of guessing the correct answers. There are computer applications available that can efficiently and cheaply accomplish this goal. Currently only the order of items on each form is randomized.

- It is important that the reliability of the renewal tests be increased. This can be done most easily by increasing the length of the renewal tests. It is therefore recommended that renewal applicants be given all 36 items on the DL 5 instead of only the first 18. It is estimated, using the Spearman-Brown Prophecy Formula, that doubling the test length would increase the overall English DL 5 renewal test internal-consistency reliability from .55 to .71, and the overall Spanish DL 5 renewal test reliability from .68 to .81. The test reliabilities would also be improved by correcting problem items, especially those with a low item-total correlation.

- Steps should be taken to ensure that field office personnel are administering only current versions of the tests in accordance with the department procedures stated in the Driver License Manual. Doing so would increase the effectiveness of randomizing the English DL 5 test every 3 months to reduce the possibility of cheating.

- The department should continue investigating the possible use of computer technology for generating the test forms and possibly creating a unique test for each applicant from a large item pool database.

- The department should collect a larger sample of Spanish language tests during the next statewide evaluation, so that item statistics can be calculated for both language groups.

- The department should consider the following strategies for increasing the level of knowledge of the applicants and thereby improving test scores:

    - Prepare and distribute information brochures and press releases that publicize the knowledge content areas and principles that applicants most frequently have problems with and emphasize the importance of reading the driver handbook before taking the test.

    - Develop and implement procedures that would better identify applicants who need to take an oral rather than written knowledge test. The oral test aids illiterate applicants who have the necessary knowledge of safe driving practices and rules of the road, but have difficulty reading the test.

    - Require waiting periods between test attempts, which is already being done for provisional license applicants. This would increase the incentive of applicants to read and study the handbook more thoroughly before their first and subsequent test attempts.

    - The department should stop the policy of returning graded test forms to encourage applicants to read the manual, rather than relying only on reviewing the old test forms to prepare for the next test.

APPENDICES

APPENDIX A

Item Statistics for English DL 5 (Rev. 2/01)
for First-Attempt Applicants by Field Office

Table A

Number of Tests (*n*) and Fail Rates for English DL 5 (Rev. 02/01)
Original and Renewal Applicants (Over All Attempts) by Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate	*n*	Fail rate
501 Sacramento	43	74.4	41	36.6
502 Los Angeles	68	69.1	17	47.1
503 San Francisco	119	37.0	59	33.9
504 Oakland	45	60.0	41	34.1
505 Fresno	18	77.8	11	45.5
506 San Diego	41	31.7	46	28.3
507 Long Beach	55	65.5	57	29.8
508 Hollywood	73	50.7	45	51.1
509 Pasadena	55	43.6	63	30.2
510 Glendale	54	57.4	45	37.8
511 Montebello	70	74.3	50	54.0
512 San Bernardino	38	68.4	30	30.0
513 Truckee	5	0.0	7	14.3
514 Culver City	50	52.0	31	25.8
515 Van Nuys	52	59.6	37	29.7
516 San Jose	62	61.3	17	41.2
517 Stockton	28	75.0	31	41.9
518 Mountain View	14	35.7	5	40.0
519 San Diego-Clairmont	74	23.0	78	29.5
520 Chico	7	0.0	18	16.7
521 Jackson	1	100.0	2	50.0
522 Oroville	5	40.0	3	0.0
523 Concord	20	45.0	38	26.3
524 Crescent City	2	50.0	2	50.0
525 Placerville	3	33.3	9	11.1
526 Eureka	10	60.0	13	15.4
527 El Centro	19	52.6	8	50.0
528 Blythe	5	60.0	3	66.7
529 Bakersfield	24	70.8	23	17.4
530 Lakeport	7	28.6	14	28.6
531 Susanville	3	0.0	5	0.0
532 Pomona	71	64.8	78	48.7
533 Madera	15	80.0	7	28.6
534 Corte Madera	18	27.8	21	23.8
535 Ukiah	4	50.0	5	20.0
536 Merced	17	70.6	18	38.9
537 Alturas	0	*	2	50.0
538 South Lake Tahoe	4	50.0	3	33.3
539 Salinas	15	46.7	17	58.8
540 Napa	8	37.5	12	25.0
541 Grass Valley	2	0.0	12	41.7
542 Santa Ana	33	45.5	28	21.4
543 Roseville	32	46.9	29	31.0
544 Quincy	6	50.0	3	0.0
545 Riverside	28	71.4	23	47.8
546 Hollister	4	100.0	12	33.3
547 San Luis Obispo	21	28.6	24	20.8

Table A (continued)

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate	*n*	Fail rate
548 Redwood City	13	15.4	19	15.8
549 Santa Barbara	16	37.5	9	22.2
550 Capitola	21	52.4	21	19.0
551 Redding	6	50.0	19	10.5
552 Yreka	1	0.0	4	75.0
553 Tulelake	0	*	0	*
554 Vallejo	22	77.3	17	35.3
555 Santa Rosa	30	50.0	36	44.4
556 El Cerrito	45	44.4	39	38.5
557 Modesto	29	62.1	22	36.4
558 Red Bluff	2	100.0	6	0.0
559 Visalia	15	53.3	27	25.9
560 Ventura	17	47.1	8	25.0
561 Woodland	4	50.0	11	63.6
562 Yuba City	25	60.0	13	46.2
563 Santa Maria	5	60.0	17	35.3
564 Colusa	1	0.0	0	*
565 Hanford	18	50.0	20	25.0
566 Mariposa	1	0.0	3	33.3
567 Seaside	14	64.3	18	27.8
568 San Andreas	5	20.0	10	20.0
569 Sonora	2	0.0	14	21.4
570 Auburn	5	60.0	15	13.3
571 Willows	0	*	2	100.0
572 Weaverville	1	0.0	0	*
573 Porterville	12	66.7	15	26.7
574 Paso Robles	6	66.7	15	26.7
575 Taft	2	0.0	4	25.0
576 Bell Gardens	49	73.5	21	52.4
577 Ridgecrest	5	40.0	2	0.0
578 Indio	21	61.9	24	45.8
579 Hayward	56	55.4	50	44.0
580 Clovis	16	62.5	27	25.9
581 Compton	53	73.6	27	59.3
582 Barstow	6	16.7	14	42.9
583 Watsonville	5	80.0	16	75.0
584 Needles	1	0.0	0	*
585 Bishop	3	66.7	4	25.0
586 Norco	28	71.4	38	26.3
587 Arleta	44	63.6	28	50.0
588 Vacaville	9	77.8	8	37.5
589 Lompoc	15	40.0	10	30.0
590 Fort Bragg	0	*	0	*
591 Whittier	0	*	0	*
592 Pittsburg	0	*	2	100.0
593 San Mateo	38	36.8	42	33.3
594 Tulare	12	66.7	8	75.0
595 Lancaster	18	72.2	26	38.5
596 Oceanside	51	45.1	51	39.2
597 Brawley	5	60.0	6	50.0
598 Davis	9	55.6	4	0.0

Table A (continued)

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate	n	Fail rate
599 Daly City	60	78.3	35	45.7
601 Paradise	2	0.0	8	12.5
602 Sacramento-South	0	*	0	*
603 Coalinga	2	50.0	3	33.3
604 Oakland Coliseum	34	64.7	32	62.5
605 Laguna Hills	23	52.2	58	25.9
606 Bellflower	56	55.4	54	40.7
607 Fullerton	61	54.1	95	38.9
608 Torrance	65	61.5	38	31.6
609 Hawthorne	39	61.5	31	58.1
610 Inglewood	63	85.7	54	68.5
611 Westminister	49	38.8	71	39.4
612 Rancho Cucamonga	0	*	0	*
613 Chula Vista	65	63.1	49	40.8
614 Spring Valley	0	*	4	50.0
615 Delano	12	83.3	3	66.7
616 Santa Monica	36	41.7	43	37.2
617 Lincoln Park	41	63.4	28	71.4
618 West Covina	78	59.0	71	39.4
619 San Pedro	35	62.9	42	31.0
620 Escondido	18	27.8	45	20.0
621 Fairfield	17	41.2	21	38.1
622 Lodi	9	55.6	25	40.0
623 Gilroy	0	*	0	*
624 Walnut Creek	35	42.9	31	19.4
625 Carmichael	30	76.7	48	33.3
626 Redlands	36	63.9	30	30.0
627 Garberville	1	0.0	1	0.0
628 Costa Mesa	36	47.2	50	30.0
629 Victorville	28	67.9	49	24.5
630 Santa Paula	9	55.6	9	33.3
631 Pleasanton	35	40.0	21	19.0
632 Santa Clara	81	43.2	49	49.0
633 Reedley	20	65.0	13	23.1
634 Petaluma	23	39.1	32	31.3
635 Hemet	8	62.5	45	40.0
636 Oxnard	23	65.2	18	38.9
637 Winnetka	53	60.4	47	31.9
638 Twentynine Palms	13	76.9	6	66.7
639 Mount Shasta	0	*	2	0.0
640 Los Gatos	22	50.0	22	50.0
641 Banning	6	66.7	16	12.5
642 Tracy	3	33.3	10	40.0
643 Fall River Mills	0	*	2	50.0
644 Fremont	57	45.6	33	33.3
645 Orland	2	0.0	3	33.3
646 Fresno-North	57	64.9	41	26.8
647 King City	5	80.0	4	0.0
648 San Clemente	20	35.0	41	22.0
649 Turlock	12	83.3	34	20.6
650 Los Banos	6	100.0	6	50.0

Table A (continued)

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate	*n*	Fail rate
655 Folsom	19	31.6	18	33.3
656 Riverside-East	20	45.0	25	44.0
657 Fontana	38	60.5	23	43.5
658 Manteca	8	62.5	11	27.3
659 Palm Springs	10	60.0	24	37.5
660 Shafter	5	60.0	5	0.0
661 Arvin	2	100.0	7	28.6
662 Newhall	11	81.8	34	17.6
663 Thousand Oaks	18	44.4	27	25.9
668 Santa Teresa	34	64.7	34	50.0
669 El Cajon	46	60.9	47	34.0
670 Goleta	8	50.0	22	27.3
672 Temecula	24	62.5	33	30.3
673 Rocklin	12	58.3	13	23.1
676 Poway	24	20.8	27	11.1
677 San Ysidro	25	48.0	7	14.3
679 Bakersfield-Southwest	21	66.7	15	26.7
680 Simi Valley	22	54.5	21	33.3
686 Novato	4	75.0	11	27.3
687 Lake Isabella	2	0.0	2	50.0
690 Palmdale	15	66.7	12	41.7
Total	3,864	55.9	3,826	35.3

Note. Office fail rates that are based on fewer than 20 test forms are likely to be unreliable, and should *not* be interpreted as accurate estimates. Fail rates are not presented by office for Spanish DL 5 originals and renewals, or for English DL 5T provisionals, because too few forms were collected to compute accurate estimates for the majority of the offices. An asterisk (*) entry indicates that the fail rate could not be computed because no test forms of that type were received. The Rancho Cucamonga field office was the only office that failed to send in any tests because it was closed during the data collection period.

[a]The figures presented for total fail rates are weighted averages.

Appendix B

Item Statistics for English DL 5 (Rev. 2/01) for First-Attempt Original Applicants

Table B1

Percentage of Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 (Rev. 2/01) Test for First-Attempt Original Applicants

Item	Answer choice	Form 1 (n = 210)	Form 2 (n = 190)	Form 3 (n = 206)	Form 4 (n = 214)	Form 5 (n = 207)	Form 6 (n = 201)	Form 7 (n = 188)	Form 8 (n = 208)	Form 9 (n = 172)	Form 10 (n = 209)
1	a	66.7	77.9	1.0	1.4	15.0	2.0	1.1	12.0	67.8	1.0
	b	13.8	8.4	4.9	7.5	66.2	95.5	94.7	77.9	13.5	9.1
	c	19.5	13.7	94.2	91.1	18.8	2.5	4.3	10.1	18.7	90.0
2	a	93.3	76.2	1.9	10.7	24.2	32.8	77.1	17.9	81.4	65.1
	b	1.4	20.6	90.3	79.4	3.9	59.2	1.1	70.5	12.8	8.6
	c	5.3	3.2	7.8	9.8	72.0	8.0	21.8	11.6	5.8	26.3
3	a	26.2	69.5	7.8	6.5	84.1	9.5	29.3	70.5	86.6	76.2
	b	71.4	21.1	5.3	68.7	1.4	8.5	8.5	10.1	4.7	18.0
	c	2.4	9.5	86.9	24.8	14.5	82.1	62.2	19.3	8.7	5.8
4	a	6.7	5.8	87.4	4.7	8.7	2.0	3.7	93.3	3.5	0.0
	b	90.4	85.3	6.3	15.9	75.8	4.5	96.3	3.8	84.3	98.6
	c	2.9	8.9	6.3	79.4	15.5	93.5	0.0	2.9	12.2	1.4
5	a	0.5	4.2	95.1	87.9	75.1	19.1	6.9	17.4	2.3	97.6
	b	6.7	85.8	1.9	6.5	4.4	2.5	78.2	66.7	89.5	1.0
	c	92.9	10.0	2.9	5.6	20.5	78.4	14.9	15.9	8.1	1.4
6	a	81.0	60.5	6.8	92.5	1.0	3.5	3.7	9.1	2.3	11.5
	b	5.7	7.4	86.8	1.9	7.7	24.4	2.1	6.3	91.3	68.9
	c	13.3	32.1	6.3	5.6	91.3	72.1	94.1	84.6	6.4	19.6
7	a	1.9	46.6	85.9	7.5	64.1	94.5	5.9	83.6	80.2	92.3
	b	1.0	34.9	4.9	85.5	13.1	3.5	78.2	12.1	8.1	1.4
	c	97.1	18.5	9.3	7.0	22.8	2.0	16.0	4.3	11.6	6.2
8	a	1.4	6.3	3.9	0.9	49.3	79.5	71.1	77.9	7.6	9.6
	b	97.6	71.6	37.3	95.3	46.3	7.0	21.9	19.2	6.4	85.6
	c	1.0	22.1	58.8	3.8	4.4	13.5	7.0	2.9	86.0	4.8
9	a	21.0	13.2	5.8	90.1	1.9	4.0	62.8	79.8	82.6	99.0
	b	66.7	74.6	7.3	5.2	2.4	11.6	27.7	8.7	6.4	0.0
	c	12.4	12.2	86.9	4.7	95.7	84.4	9.6	11.5	11.0	1.0
10	a	1.4	94.7	66.8	91.6	27.5	25.9	5.3	11.2	5.2	23.1
	b	15.7	2.6	25.4	6.5	7.7	7.0	0.0	9.7	94.2	63.9
	c	82.9	2.6	7.8	1.9	64.7	67.2	94.7	79.1	0.6	13.0
11	a	3.3	15.9	2.9	9.8	11.6	12.5	3.2	30.4	77.1	79.8
	b	11.5	5.3	86.4	86.4	3.4	14.0	3.2	66.2	5.3	7.7
	c	85.2	78.8	10.7	3.7	85.0	73.5	93.6	3.4	17.6	12.5
12	a	14.8	12.6	9.2	1.4	20.0	86.1	8.5	4.3	82.6	53.6
	b	65.6	76.8	19.9	95.8	9.3	0.0	88.8	84.1	14.0	7.2
	c	19.6	10.5	70.9	2.8	70.7	13.9	2.7	11.6	3.5	39.2
13	a	77.0	14.7	2.0	65.4	73.0	72.5	5.3	11.8	53.5	4.8
	b	13.4	5.8	95.6	11.7	21.6	20.5	9.6	80.4	32.0	86.1
	c	9.6	79.5	2.4	22.9	5.4	7.0	85.1	7.8	14.5	9.1
14	a	3.3	12.8	16.1	63.6	12.6	10.1	0.0	95.7	2.3	11.5
	b	1.9	15.4	82.9	17.8	10.6	1.0	100.0	1.4	80.8	76.1
	c	94.7	71.8	1.0	18.7	76.8	88.9	0.0	2.9	16.9	12.4
15	a	0.0	68.8	89.8	7.5	0.0	7.0	3.2	85.0	63.5	21.7
	b	61.7	20.1	5.8	5.6	0.5	85.6	86.7	5.3	25.3	20.3
	c	38.3	11.1	4.4	86.9	99.5	7.5	10.1	9.7	11.2	58.0
16	a	98.6	2.6	12.1	6.5	10.7	58.8	89.4	50.0	0.6	64.1
	b	0.0	91.1	76.7	65.0	85.4	19.6	7.4	7.8	94.8	17.2
	c	1.4	6.3	11.2	28.5	3.9	21.6	3.2	42.2	4.7	18.7
17	a	13.3	93.7	23.8	19.2	60.2	78.6	22.9	11.6	3.5	90.4
	b	75.2	3.7	1.0	72.9	9.7	10.9	3.7	73.9	5.8	2.9
	c	11.4	2.6	75.2	7.9	30.1	10.4	73.4	14.5	90.7	6.7
18	a	59.0	6.8	7.3	5.1	91.3	0.0	85.6	34.3	11.6	1.9
	b	14.8	8.4	90.7	7.5	3.9	6.0	12.8	57.0	78.5	9.1
	c	26.2	84.7	2.0	87.4	4.8	94.0	1.6	8.7	9.9	89.0
19	a	89.0	3.2	5.4	4.7	2.9	7.5	97.9	5.8	89.0	4.8
	b	2.4	85.1	91.2	24.4	84.0	11.4	0.0	4.8	1.7	91.4
	c	8.6	11.7	3.4	70.9	13.1	81.1	2.1	89.4	9.3	3.8

Table B1 (continued)

Item	Answer choice	Form 1 (*n* = 210)	Form 2 (*n* = 190)	Form 3 (*n* = 206)	Form 4 (*n* = 214)	Form 5 (*n* = 207)	Form 6 (*n* = 201)	Form 7 (*n* = 188)	Form 8 (*n* = 208)	Form 9 (*n* = 172)	Form 10 (*n* = 209)
20	a	<u>84.8</u>	10.5	9.2	22.5	15.0	5.0	9.6	<u>90.4</u>	8.8	3.8
	b	10.5	<u>86.8</u>	<u>63.1</u>	<u>69.0</u>	<u>80.1</u>	<u>87.5</u>	11.7	3.4	<u>56.1</u>	<u>75.5</u>
	c	4.8	2.6	27.7	8.5	4.9	7.5	<u>78.7</u>	6.3	35.1	20.7
21	a	<u>87.6</u>	11.1	3.9	<u>81.7</u>	**1.4**	4.0	24.5	4.9	<u>90.0</u>	11.1
	b	**1.4**	16.3	6.3	3.3	**1.9**	8.0	28.2	<u>63.1</u>	3.5	<u>75.0</u>
	c	11.0	<u>72.6</u>	<u>89.8</u>	15.0	<u>96.6</u>	<u>88.1</u>	<u>47.3</u>	32.0	6.5	13.9
22	a	7.7	**0.5**	13.7	3.3	24.6	34.8	**1.6**	16.8	2.3	<u>92.8</u>
	b	<u>87.6</u>	<u>86.2</u>	<u>68.6</u>	3.3	<u>55.6</u>	16.4	3.7	13.9	**1.2**	5.7
	c	4.8	13.2	17.6	<u>93.5</u>	19.8	<u>48.8</u>	<u>94.7</u>	<u>69.2</u>	<u>96.5</u>	**1.4**
23	a	6.2	3.7	20.0	**1.9**	**0.0**	4.0	11.7	5.3	5.8	<u>80.9</u>
	b	<u>85.2</u>	19.5	8.3	2.3	18.8	27.5	<u>85.1</u>	15.0	7.6	11.0
	c	8.6	<u>76.8</u>	<u>71.7</u>	<u>95.8</u>	<u>81.2</u>	<u>68.5</u>	3.2	<u>79.6</u>	<u>86.5</u>	8.1
24	a	<u>79.9</u>	<u>84.7</u>	2.9	<u>82.7</u>	4.4	5.5	<u>41.7</u>	21.7	**1.2**	15.0
	b	12.0	7.4	12.6	3.3	16.5	**0.5**	2.1	<u>74.9</u>	**1.7**	6.8
	c	8.1	7.9	<u>84.5</u>	14.0	<u>79.1</u>	<u>94.0</u>	**56.1**	3.4	<u>97.1</u>	<u>78.3</u>
25	a	<u>91.9</u>	21.2	<u>73.8</u>	2.3	<u>58.7</u>	<u>95.0</u>	16.0	8.8	3.5	14.4
	b	4.8	16.9	21.4	<u>92.5</u>	23.8	5.0	<u>82.9</u>	13.2	<u>70.8</u>	<u>80.8</u>
	c	3.3	<u>61.9</u>	4.9	5.2	17.5	**0.0**	**1.1**	<u>78.0</u>	25.7	4.8
26	a	6.2	5.8	**1.9**	<u>68.7</u>	7.8	4.0	5.3	<u>86.1</u>	4.1	<u>78.8</u>
	b	<u>90.4</u>	<u>78.4</u>	<u>84.0</u>	7.5	7.3	5.0	**1.6**	7.7	<u>82.8</u>	16.3
	c	3.3	15.8	14.1	23.8	<u>85.0</u>	<u>91.0</u>	<u>93.1</u>	6.3	13.0	4.8
27	a	<u>90.4</u>	2.6	**0.5**	<u>90.1</u>	9.7	15.0	11.8	<u>87.3</u>	17.4	8.1
	b	5.7	2.1	**1.5**	**1.9**	<u>80.2</u>	18.5	11.8	8.8	7.6	10.0
	c	3.8	<u>95.3</u>	<u>98.1</u>	8.0	10.1	<u>66.5</u>	<u>76.5</u>	3.9	<u>75.0</u>	<u>81.8</u>
28	a	16.7	5.3	19.4	8.4	27.5	5.5	<u>81.7</u>	<u>87.4</u>	7.0	<u>44.5</u>
	b	10.0	8.4	<u>78.2</u>	<u>59.3</u>	<u>51.7</u>	<u>47.0</u>	8.6	7.7	<u>82.0</u>	6.2
	c	<u>73.3</u>	<u>86.3</u>	2.4	32.2	20.8	**47.5**	9.7	4.8	11.0	**49.3**
29	a	7.2	<u>72.1</u>	2.9	7.9	5.3	16.4	7.5	3.4	5.8	9.2
	b	5.3	4.7	6.3	<u>62.6</u>	**1.4**	<u>62.7</u>	<u>80.2</u>	<u>80.3</u>	<u>88.9</u>	<u>88.8</u>
	c	<u>87.6</u>	23.2	<u>90.8</u>	29.4	<u>93.2</u>	20.9	12.3	16.3	5.3	**1.9**
30	a	3.3	<u>92.1</u>	29.1	**1.4**	<u>89.8</u>	**2.0**	14.4	<u>62.0</u>	30.4	**58.2**
	b	5.2	6.3	7.3	**0.9**	8.7	<u>91.5</u>	<u>82.4</u>	9.3	<u>54.4</u>	<u>41.3</u>
	c	<u>91.4</u>	**1.6**	<u>63.6</u>	<u>97.7</u>	**1.5**	6.5	3.2	28.8	15.2	**0.5**
31	a	<u>89.0</u>	**1.1**	19.9	14.6	5.3	<u>81.0</u>	4.3	2.9	11.0	8.6
	b	9.6	<u>95.3</u>	5.8	<u>67.1</u>	<u>91.3</u>	3.0	<u>89.9</u>	<u>85.0</u>	<u>73.8</u>	<u>77.0</u>
	c	**1.4**	3.7	<u>74.3</u>	18.3	3.4	16.0	5.9	12.1	15.1	14.4
32	a	<u>87.1</u>	2.1	3.4	9.5	14.0	16.9	**1.6**	<u>88.9</u>	<u>87.8</u>	12.4
	b	**1.4**	6.8	<u>54.4</u>	<u>79.6</u>	9.7	<u>64.2</u>	<u>81.8</u>	4.3	4.7	22.5
	c	11.4	<u>91.1</u>	42.2	10.9	<u>76.3</u>	18.9	16.6	6.7	7.6	<u>65.1</u>
33	a	8.1	8.5	13.7	<u>65.7</u>	3.4	16.4	39.9	8.2	3.5	**1.0**
	b	19.0	5.3	<u>73.7</u>	13.6	36.1	**2.0**	<u>52.7</u>	29.8	14.0	12.0
	c	<u>72.9</u>	<u>86.2</u>	12.7	20.7	<u>60.5</u>	<u>81.6</u>	7.4	<u>62.0</u>	<u>82.6</u>	<u>87.1</u>
34	a	19.5	5.3	<u>81.6</u>	**1.9**	10.1	28.4	<u>81.8</u>	5.3	18.8	**1.0**
	b	11.9	<u>77.4</u>	14.1	<u>96.7</u>	3.9	<u>65.2</u>	3.7	<u>94.2</u>	<u>58.2</u>	3.8
	c	<u>68.6</u>	17.4	4.4	**1.4**	<u>86.0</u>	6.5	14.4	**0.5**	22.9	<u>95.2</u>
35	a	2.9	6.9	**1.5**	<u>92.5</u>	2.4	<u>88.1</u>	10.1	<u>93.8</u>	<u>91.9</u>	<u>70.8</u>
	b	21.5	<u>71.3</u>	2.9	3.3	<u>94.7</u>	8.0	9.6	**1.4**	2.9	3.3
	c	<u>75.6</u>	21.8	<u>95.6</u>	4.2	2.9	4.0	<u>80.3</u>	4.8	5.2	25.8
36	a	25.2	21.6	9.2	37.4	21.7	<u>77.1</u>	<u>68.6</u>	5.8	4.7	15.9
	b	<u>57.1</u>	2.6	<u>79.6</u>	7.9	**0.5**	16.4	13.3	9.6	<u>93.0</u>	<u>71.0</u>
	c	17.6	<u>75.8</u>	11.2	<u>54.7</u>	<u>77.8</u>	6.5	18.1	<u>84.6</u>	2.3	13.0

Note. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distracter selection rate is too low or too high.

Table B2

Item-Total Correlation for Each Item on Each Form of the English
DL 5 (Rev. 2/01) Test for First-Attempt Original Applicants

Item	Form 1 ($n = 210$)	Form 2 ($n = 190$)	Form 3 ($n = 206$)	Form 4 ($n = 214$)	Form 5 ($n = 207$)	Form 6 ($n = 201$)	Form 7 ($n = 188$)	Form 8 ($n = 208$)	Form 9 ($n = 172$)	Form 10 ($n = 209$)
1	.38	.12	.18	.10	.40	.05	.15	.17	.31	.28
2	.38	.27	.13	.05	.00	.27	.16	.39	.47	.18
3	.23	.14	.09	.41	.23	.37	.42	.33	.32	.18
4	.41	.27	.36	.18	.14	.30	.25	.27	.25	.28
5	.16	.29	.07	.35	.29	.30	.08	.06	.40	.07
6	.27	.43	.29	.30	.21	.54	.07	.18	.20	.21
7	.18	.05	.04	.22	.35	.33	.34	.35	.29	.32
8	.29	.39	.23	.16	.00	.09	.38	.42	.25	.43
9	.21	.24	.37	.34	.22	.22	.27	.20	.22	-.05
10	.18	.24	.17	.24	.28	.46	.19	.26	.20	.41
11	.27	.33	.31	.18	.33	.34	.25	-.02	.33	.16
12	.21	.33	.17	.27	.30	.31	.37	.39	.39	.31
13	.31	.37	.39	.20	.37	.37	.45	.30	.33	.37
14	.34	-.01	.41	.44	.42	.45	*	.18	.34	.32
15	.34	.26	.25	.33	.03	.28	.27	.28	.24	.20
16	.12	.21	-.01	.45	.30	.11	.37	.34	.14	.33
17	.28	.24	.04	.32	.38	.34	.22	.47	.25	.16
18	.15	.38	.42	.36	.41	.37	.28	.01	.37	.24
19	.42	.11	.32	.28	.35	.45	.10	.35	.18	.25
20	.35	.20	.23	.24	.35	.08	.37	.46	.56	.41
21	.06	.32	.28	.41	.38	.42	.23	.17	.37	.24
22	.54	.31	.26	.27	.34	.14	.19	.17	.01	.17
23	.39	.15	.02	.24	.32	.50	.15	.37	.26	.13
24	.31	.45	.40	.40	.27	.34	.27	.29	.17	.35
25	.23	-.01	.07	.26	.26	.39	.28	.43	.13	.25
26	.38	.35	.31	.36	.20	.41	.22	.35	.54	.28
27	.38	.15	-.02	.20	.01	.45	.31	.26	.44	.28
28	.38	.28	.34	.22	.22	.14	.36	.32	.29	.18
29	.06	.25	.44	.28	.33	.02	.15	.33	.30	.34
30	.43	.17	.14	.24	.32	.29	.38	.12	.31	.23
31	.30	.22	.10	.43	.42	.43	.14	.08	.52	.34
32	.27	.42	.20	.33	.21	.33	.23	.44	.33	.25
33	.27	.23	.17	.06	.22	.10	.15	.24	.25	.09
34	-.01	.23	.44	.22	.24	.43	.18	.19	.29	.26
35	.28	.24	.20	.24	.39	.42	.20	.16	.18	.25
36	.07	.46	.37	.18	.37	.37	.30	.41	.36	.31

Note. Shading indicates that an item needs to be reviewed and possibly revised or replaced because the item-total correlation is negative or less than .10. An asterisk (*) entry indicates that the item-total correlation could not be calculated because the item had insufficient variance.

Table B3

Percentage of First-Attempt Original Applicants Who Would Pass if Different Cut-Points (Number Missed) Were Used for Each Form of the English DL 5 (Rev. 2/01)

Number Missed	Form 1 (n = 210)	Form 2 (n = 190)	Form 3 (n = 206)	Form 4 (n = 214)	Form 5 (n = 207)	Form 6 (n = 201)	Form 7 (n = 188)	Form 8 (n = 208)	Form 9 (n = 172)	Form 10 (n = 209)	Total (N = 2,005)
0	2.4	0.0	1.9	3.7	2.4	1.5	2.7	1.4	6.4	2.4	2.4
1	8.6	6.3	3.9	9.3	6.3	6.0	8.0	3.4	11.6	5.7	6.8
2	16.7	13.2	11.7	15.4	9.7	14.9	14.4	7.7	18.6	11.0	13.2
3	28.1	19.5	19.4	22.9	15.5	23.4	23.9	17.3	29.1	14.8	21.2
4	35.7	28.9	31.6	34.1	23.2	29.4	33.0	26.9	37.2	24.9	30.4
5	47.1	39.5	42.7	44.4	32.4	36.8	43.6	34.1	49.4	32.1	40.0
6	56.2	48.4	53.4	55.1	41.1	47.8	53.2	44.2	57.6	40.7	49.6
7	67.1	56.3	65.0	60.7	53.1	56.2	62.2	56.7	62.8	50.2	59.0
8	73.8	63.2	72.3	67.3	61.8	62.7	67.6	63.9	68.6	55.5	65.6
9	79.5	71.6	79.6	73.8	68.1	67.7	73.4	71.2	73.3	65.1	72.3
10	82.9	78.4	85.9	79.4	72.9	74.1	78.7	77.9	77.9	69.4	77.8
11	86.2	82.6	88.8	83.6	77.8	78.6	81.9	82.2	79.1	77.0	81.8
12	89.0	87.9	91.7	86.9	81.6	83.1	85.1	86.5	86.0	80.9	85.9
13	90.0	91.1	92.7	89.3	86.0	85.6	92.0	88.0	89.0	87.6	89.1
14	91.9	93.2	94.2	92.1	87.9	86.1	94.1	89.9	90.7	90.0	91.0
15	94.8	94.7	94.7	94.4	90.8	89.1	95.7	93.7	93.0	92.3	93.3
16+	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. The shaded line highlights the pass rates at the current passing standard of six allowable errors.

Table B4

Summary of Problem Items on Each Form of the English
DL 5 (Rev. 2/01) Test for First-Attempt Original Applicants

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items	Form 6 items	Form 7 items	Form 8 items	Form 9 items	Form 10 items
Item-total correlation too low or negative[a]	21, 29, 34, 36	7, 14, 25	3, 5, 7, 16, 17, 23, 25, 27	2, 33	2, 8, 15, 27	1, 8, 20, 29	5, 6, 14	5, 11, 18, 31	22	5, 9, 33
Pass rate too high[b]	7, 8, 16	27, 31	5, 13, 27, 35	8, 12, 23, 30, 34	9, 15, 21	1	4, 14, 19	14	22, 24	4, 5, 9, 34
Pass rate too low[c]	18, 36	7	8, 32	28, 36	8, 22, 25, 28	2, 16, 22, 28	21, 24, 33	16, 18	13, 20, 30, 34	12, 15, 28, 30
Pass rate too high or low and item-total correlation too low or negative	36	7	5, 27		8, 15	1	14	18	22	5, 9
Distracter selected too often[d]					8	28	24	16		28, 30
Distracter selected too infrequently[e]	2, 5, 7, 8, 10, 14, 15, 16, 21, 31, 32	22, 30, 31	1, 2, 5, 13, 14, 17, 18, 26, 27, 35	1, 6, 8, 10, 12, 23, 27, 30, 34	3, 6, 9, 15, 21, 23, 29, 30, 36	1, 4, 7, 12, 14, 18, 24, 25, 30, 33	1, 2, 4, 10, 14, 18, 19, 22, 25, 26, 32	14, 34, 35	10, 16, 19, 22, 24	1, 4, 5, 7, 9, 18, 22, 29, 30, 33, 34

[a]The item-total correlation was negative or less than .10. [b]More than 95% of applicants answered the item correctly. [c]Less than 60% of applicants answered the item correctly. [d]A distracter was chosen more, or almost, as often as the correct answer to the item. [e]A distracter was selected by 2% or fewer applicants.

Appendix C

Item Statistics for English DL 5 (Rev. 2/01) for First-Attempt Renewal Applicants

Table C1

Percentage of Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 (Rev. 2/01) Test for First-Attempt Renewal Applicants

Item	Answer choice	Form 1 (n = 305)	Form 2 (n = 314)	Form 3 (n = 264)	Form 4 (n = 262)	Form 5 (n = 274)	Form 6 (n = 259)	Form 7 (n = 281)	Form 8 (n = 288)	Form 9 (n = 306)	Form 10 (n = 278)
1	a	76.7	82.4	1.5	3.8	4.7	1.5	0.7	17.1	75.2	1.4
	b	7.2	4.5	3.8	6.9	84.3	97.3	96.4	71.1	2.6	6.8
	c	16.1	13.1	94.7	89.3	10.9	1.2	2.8	11.8	22.2	91.7
2	a	98.0	88.9	1.9	6.1	35.5	21.6	78.1	8.7	89.1	66.2
	b	0.7	10.2	81.8	84.3	3.7	72.6	1.1	85.4	3.9	12.2
	c	1.3	1.0	16.3	9.6	60.8	5.8	20.9	5.9	6.9	21.6
3	a	18.4	62.0	8.3	2.7	79.6	6.2	21.8	66.0	94.8	81.5
	b	81.6	29.4	4.5	84.4	2.2	4.6	1.8	23.3	0.0	13.1
	c	0.0	8.6	87.1	13.0	18.2	89.2	76.4	10.8	5.2	5.5
4	a	3.0	1.0	91.6	2.3	9.9	0.8	2.5	92.3	1.3	1.4
	b	94.4	93.3	4.9	21.5	76.8	2.7	96.8	3.1	92.1	97.5
	c	2.6	5.7	3.4	76.2	13.2	96.5	0.7	4.5	6.6	1.1
5	a	0.0	2.9	96.6	93.9	79.1	14.5	10.8	15.0	2.3	97.5
	b	3.6	92.3	2.3	3.4	2.6	0.8	77.4	67.6	92.5	0.7
	c	96.4	4.8	1.1	2.7	18.3	84.8	11.8	17.4	5.2	1.8
6	a	82.1	73.7	6.8	95.4	0.7	2.7	1.4	7.3	2.9	8.4
	b	5.0	6.1	91.7	1.1	5.1	15.1	1.8	5.2	93.1	77.4
	c	12.9	20.2	1.5	3.4	94.2	82.2	96.8	87.5	3.9	14.2
7	a	1.0	49.7	85.2	4.6	71.0	97.7	6.8	87.2	89.1	97.5
	b	0.7	36.2	4.2	90.8	7.7	1.6	83.2	10.1	6.3	0.4
	c	98.3	14.1	10.6	4.6	21.3	0.8	10.0	2.8	4.6	2.2
8	a	0.7	5.8	0.4	1.1	52.4	86.4	82.9	84.7	11.1	6.8
	b	99.3	77.3	34.1	95.8	44.3	5.1	9.3	9.7	4.9	91.0
	c	0.0	16.9	65.5	3.1	3.3	8.6	7.8	5.6	84.0	2.2
9	a	24.0	10.8	1.5	94.6	1.1	5.8	74.4	81.5	87.2	96.0
	b	71.1	81.5	4.5	2.7	2.6	12.0	23.1	7.7	3.6	0.7
	c	4.9	7.6	93.9	2.7	96.4	82.2	2.5	10.8	9.2	3.2
10	a	1.3	97.1	74.9	98.1	26.0	19.1	3.2	14.8	3.0	15.2
	b	12.8	1.6	21.3	1.5	5.1	3.5	0.7	12.7	96.1	78.0
	c	85.9	1.3	3.8	0.4	68.9	77.4	96.1	72.5	1.0	6.9
11	a	2.0	7.3	0.8	8.8	7.3	11.7	1.4	27.9	82.0	85.6
	b	5.3	1.6	95.1	89.7	1.8	6.6	2.5	68.6	3.6	5.8
	c	92.8	91.1	4.2	1.5	90.8	81.7	96.1	3.5	14.4	8.7
12	a	12.2	4.1	2.3	1.1	5.8	90.3	2.5	9.4	83.3	49.8
	b	66.3	90.4	15.6	97.7	6.6	0.0	95.4	78.3	13.1	4.7
	c	21.5	5.4	82.1	1.1	87.6	9.7	2.1	12.2	3.6	45.5
13	a	80.9	14.7	1.9	51.4	70.1	78.6	6.8	13.3	61.3	2.9
	b	7.6	6.7	96.2	16.6	24.8	15.6	7.2	82.1	28.9	95.3
	c	11.6	78.6	1.9	32.0	5.1	5.8	86.0	4.6	9.8	1.8
14	a	2.6	13.9	11.0	71.6	4.4	7.4	0.0	97.2	2.0	7.6
	b	1.6	10.0	89.0	12.6	6.2	0.8	99.6	0.7	92.1	88.0
	c	95.7	76.1	0.0	15.7	89.4	91.9	0.4	2.1	5.9	4.3
15	a	0.7	64.4	92.8	6.1	0.4	7.8	2.8	85.1	78.7	15.8
	b	67.5	28.5	4.9	3.8	0.4	88.3	90.4	7.3	11.1	10.8
	c	31.8	7.1	2.3	90.1	99.3	3.9	6.8	7.6	10.2	73.4
16	a	98.7	6.1	10.7	1.9	13.1	73.5	92.5	54.5	2.0	68.4
	b	0.0	88.9	74.0	84.2	85.8	13.6	5.0	8.0	96.4	11.3
	c	1.3	5.1	15.3	13.8	1.1	12.8	2.5	37.4	1.6	20.4
17	a	8.9	98.7	28.4	15.0	82.5	86.9	20.3	5.6	1.6	92.1
	b	77.2	1.0	0.0	80.8	5.1	11.2	0.4	88.5	4.9	2.2
	c	13.9	0.3	71.6	4.2	12.4	1.9	79.4	5.9	93.4	5.8
18	a	56.9	1.9	5.7	2.7	97.1	0.0	92.2	33.8	6.5	1.1
	b	7.9	4.5	93.9	7.3	0.7	3.9	5.0	56.1	85.6	9.0
	c	35.2	93.6	0.4	90.1	2.2	96.1	2.8	10.1	7.8	89.9

Note. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distracter selection rate is too low or too high.

Table C2

Item-Total Correlation for Each Item on Each Form of the English
DL 5 (Rev. 2/01) Test for First-Attempt Renewal Applicants

Item	Form 1 ($n = 305$)	Form 2 ($n = 314$)	Form 3 ($n = 264$)	Form 4 ($n = 262$)	Form 5 ($n = 274$)	Form 6 ($n = 259$)	Form 7 ($n = 281$)	Form 8 ($n = 288$)	Form 9 ($n = 306$)	Form 10 ($n = 278$)
1	.21	.14	.09	.12	.21	.23	.17	.20	.28	.38
2	.22	.24	.06	.21	.06	.06	.08	.31	.39	.13
3	.08	.16	-.08	.35	.07	.18	.29	.22	.26	.10
4	.31	.05	.20	.04	.17	.08	.43	.16	.25	.39
5	.05	.25	.01	.37	.25	.08	.15	.24	.25	.05
6	.24	.30	.23	.31	.18	.31	.26	.23	.08	.11
7	.01	.09	.14	.08	.20	.25	.31	.22	.26	.22
8	.07	.26	.23	.04	.16	.21	.19	.24	.24	.23
9	.06	.22	.29	.24	.18	.18	.10	.31	.20	.28
10	.19	.22	.14	.14	.16	.26	.11	.33	.42	.12
11	.26	.22	.38	.34	.16	.29	.15	.03	.23	.14
12	.25	.18	.20	.26	.12	.16	.24	.22	.18	.24
13	.26	.22	.16	.11	.20	-.02	.32	.30	.14	.28
14	.23	.20	.20	.34	.12	.41	.18	.08	.21	.32
15	.23	.20	.32	.25	.13	.16	.08	.22	.17	-.01
16	.19	.16	.11	.33	.20	.06	.38	.18	.09	.29
17	.18	.16	.05	.35	.15	.27	.16	.23	.22	.25
18	.02	.32	.33	.39	.16	.29	.28	.17	.24	.28

Note. Shading indicates that an item needs to be reviewed and possibly revised or replaced because the item-total correlation is negative or less than .10.

Table C3

Percentage of First-Attempt Renewal Applicants Who Would Pass if Different Cut-Points Were Used for Each Form of the English DL 5 (Rev. 2/01)

Number Missed	Form 1 ($n = 305$)	Form 2 ($n = 314$)	Form 3 ($n = 264$)	Form 4 ($n = 262$)	Form 5 ($n = 274$)	Form 6 ($n = 259$)	Form 7 ($n = 281$)	Form 8 ($n = 288$)	Form 9 ($n = 306$)	Form 10 ($n = 278$)	Total ($N = 2,831$)
0	8.2	8.9	14.0	12.2	5.5	12.0	20.3	5.6	14.1	8.3	10.8
1	27.2	22.6	35.6	40.8	19.7	34.4	42.0	17.7	41.8	24.8	30.5
2	47.5	42.0	58.3	61.8	38.7	57.1	64.8	30.6	64.1	47.1	51.0
3	69.8	60.2	73.9	77.1	53.3	75.7	83.6	49.7	78.1	69.1	68.9
4	82.6	73.6	86.7	84.0	66.8	86.5	91.5	61.1	85.0	82.0	79.8
5	90.5	85.4	92.0	91.6	84.3	93.1	94.3	72.9	91.5	91.0	88.6
6	94.1	91.1	97.3	94.7	93.1	95.8	96.8	85.1	94.8	94.6	93.6
7	96.1	95.9	98.1	96.6	97.1	97.3	97.2	91.0	97.4	97.1	96.3
8	98.7	97.8	99.2	97.7	98.2	98.8	97.9	93.4	97.7	97.5	97.7
9	99.7	99.4	99.6	98.9	98.9	99.2	99.3	96.5	99.0	98.6	98.9
10+	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. The shaded line highlights the pass rates at the current passing standard of three allowable errors.

Table C4

Summary of Problem Items on Each Form of the English DL 5 (Rev. 2/01) Test for First-Attempt Renewal Applicants

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items	Form 6 items	Form 7 items	Form 8 items	Form 9 items	Form 10 items
Item-total correlation too low or negative[a]	3, 5, 7, 8, 9, 18	4, 7	1, 2, 3, 5, 17	4, 7, 8	2, 3	2, 4, 5, 13, 16	2, 15	11, 14	6, 16	5, 15
Pass rate too high[b]	2, 5, 7, 8, 14, 16	10, 17	5, 11, 13	6, 8, 10, 12	9, 15, 18	1, 4, 7, 18	1, 4, 6, 10, 11, 12, 14	14	10, 16	4, 5, 7, 9, 13
Pass rate too low[c]	18	7		13	8			16, 18		12
Pass rate too high or low <u>and</u> item-total correlation too low or negative	5, 7, 8, 18	7	5	8		4		14	16	5
Distracter selected too often[d]					8					12
Distracter selected too infrequently[e]	2, 3, 5, 7, 8, 10, 11, 14, 15, 16	2, 4, 10, 11, 17, 18	1, 2, 5, 6, 8, 9, 11, 13, 14, 17, 18	6, 8, 10, 11, 12, 16	6, 9, 11, 15, 16, 18	1, 4, 5, 7, 12, 14, 17, 18	1, 2, 3, 4, 6, 10, 11, 14, 17	14	3, 4, 10, 14, 16, 17	1, 4, 5, 7, 9, 13, 18

[a]The item-total correlation was negative or less than .10. [b]More than 95% of applicants answered the item correctly. [c]Less than 60% of applicants answered the item correctly. [d]A distracter was chosen more, or almost, as often as the correct answer to the item. [e]A distracter was selected by 2% or fewer applicants.

43

Appendix D

Item Statistics for English DL 5T (Rev. 2/01) for First-Attempt Provisional Applicants

Table D1

Percentage of Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5T (Rev. 2/01) Test for First-Attempt Provisional Applicants

Item	Answer choice	Form 1 (n = 200	Form 2 (n = 189)	Form 3 (n = 166)	Form 4 (n = 176)	Form 5 (n = 180)
1	a	**0.5**	6.3	82.5	84.6	22.8
	b	90.0	85.7	7.2	13.7	5.0
	c	9.5	7.9	10.2	**1.7**	72.2
2	a	2.5	**0.5**	7.8	9.7	73.2
	b	3.0	**1.1**	10.2	16.5	8.4
	c	94.5	98.4	81.9	73.9	18.4
3	a	19.6	3.2	91.0	67.6	2.8
	b	9.5	73.5	6.0	**1.7**	2.8
	c	70.9	23.3	3.0	30.7	94.4
4	a	56.5	81.4	6.0	17.0	**53.9**
	b	9.5	7.4	3.0	5.1	4.4
	c	34.0	11.2	91.0	77.8	41.7
5	a	88.0	2.1	4.2	78.9	27.9
	b	7.5	95.2	6.7	13.1	60.9
	c	4.5	2.6	89.1	8.0	11.2
6	a	9.0	4.8	9.8	26.7	78.3
	b	68.0	87.8	57.9	66.5	6.7
	c	23.0	7.4	32.3	6.8	15.0
7	a	**1.0**	4.2	88.6	6.3	**1.1**
	b	96.0	12.2	6.6	76.7	2.8
	c	3.0	83.6	4.8	17.0	96.1
8	a	5.0	63.8	4.2	18.8	2.2
	b	79.0	30.3	6.6	67.0	93.9
	c	16.0	5.9	89.2	14.2	3.9
9	a	63.0	86.2	4.2	58.9	81.6
	b	24.0	2.1	27.7	30.9	3.4
	c	13.0	11.6	68.1	10.3	15.1
10	a	96.5	2.1	6.0	4.5	7.8
	b	2.5	3.2	5.4	86.9	90.0
	c	**1.0**	94.7	88.6	8.5	2.2
11	a	97.0	**0.5**	39.8	14.9	65.4
	b	**1.0**	96.3	55.4	15.5	34.1
	c	**2.0**	3.2	4.8	69.5	**0.6**
12	a	45.7	13.2	58.4	7.4	12.8
	b	11.6	82.5	19.9	90.3	84.4
	c	**42.7**	4.2	21.7	2.3	2.8
13	a	**2.0**	94.7	81.9	19.3	8.3
	b	3.5	2.6	**0.6**	75.6	75.0
	c	94.5	2.6	17.5	5.1	16.7
14	a	7.5	91.5	18.7	87.5	3.3
	b	**0.5**	**1.6**	5.4	3.4	5.6
	c	92.0	6.9	75.9	9.1	91.1
15	a	**2.0**	6.9	**1.2**	**0.0**	9.4
	b	**2.0**	88.4	2.4	96.6	82.2
	c	96.0	4.8	96.4	3.4	8.3
16	a	3.0	2.7	2.4	4.5	82.8
	b	18.0	25.5	25.9	12.5	16.1
	c	79.0	71.8	71.7	83.0	**1.1**
17	a	70.9	59.3	8.5	61.4	9.6
	b	3.5	12.7	16.4	15.3	79.2
	c	25.6	28.0	75.2	23.3	11.2
18	a	79.5	9.0	27.1	9.7	88.9
	b	6.5	81.4	53.6	23.9	7.8
	c	14.0	9.6	19.3	66.5	3.3
19	a	76.5	5.8	6.6	98.9	97.8
	b	17.0	75.7	12.0	**0.6**	**1.7**
	c	6.5	18.5	81.3	**0.6**	**0.6**

44

Table D1 (continued)

Item	Answer choice	Form 1 (*n* = 200	Form 2 (*n* = 189)	Form 3 (*n* = 166)	Form 4 (*n* = 176)	Form 5 (*n* = 180)
20	a	74.5	4.8	3.6	92.0	85.0
	b	23.0	2.1	94.0	5.1	12.2
	c	2.5	93.1	2.4	2.9	2.8
21	a	15.5	2.1	**1.8**	2.8	**0.6**
	b	78.5	4.2	97.0	79.0	96.7
	c	6.0	93.7	**1.2**	18.2	2.8
22	a	10.5	4.2	8.5	87.5	15.6
	b	6.0	86.8	78.2	9.1	6.1
	c	83.5	9.0	13.3	3.4	78.3
23	a	8.5	4.2	75.6	17.1	**1.1**
	b	82.0	20.1	2.4	12.6	10.6
	c	9.5	75.7	22.0	70.3	88.3
24	a	97.5	2.6	13.9	88.0	18.3
	b	**1.0**	5.8	64.8	10.9	70.0
	c	**1.5**	91.5	21.2	**1.1**	11.7
25	a	8.0	**1.1**	**44.8**	89.8	21.7
	b	71.5	97.9	14.5	6.8	25.0
	c	20.5	**1.1**	40.6	3.4	53.3
26	a	12.0	5.8	27.1	16.0	69.4
	b	6.0	11.1	6.0	79.4	17.2
	c	82.0	83.1	66.9	4.6	13.3
27	a	3.0	94.2	**1.2**	5.1	3.3
	b	92.5	2.1	88.5	30.7	73.9
	c	4.5	3.7	10.3	64.2	22.8
28	a	29.5	**1.6**	4.8	2.3	20.6
	b	8.5	94.1	**1.8**	10.8	78.3
	c	62.0	4.3	93.4	86.9	**1.1**
29	a	53.8	14.3	62.7	7.4	64.2
	b	25.6	68.3	10.8	90.3	28.5
	c	20.6	17.5	26.5	2.3	7.3
30	a	22.1	89.9	92.8	50.6	2.2
	b	76.4	7.4	7.2	26.7	18.9
	c	**1.5**	2.7	**0.0**	22.7	78.9
31	a	20.1	28.2	18.8	4.5	45.3
	b	21.6	71.3	5.5	79.5	10.1
	c	58.3	**0.5**	75.8	15.9	**44.7**
32	a	94.5	**0.5**	13.3	93.8	**1.7**
	b	4.0	97.4	81.8	3.4	95.0
	c	**1.5**	2.1	4.8	2.8	3.3
33	a	**2.0**	3.2	10.8	44.9	86.7
	b	88.5	**1.1**	86.1	8.0	4.4
	c	9.5	95.8	3.0	**47.2**	8.9
34	a	15.0	94.7	24.2	80.1	21.1
	b	4.0	**1.6**	75.2	12.5	24.4
	c	81.0	3.7	**0.6**	7.4	54.4
35	a	4.0	93.1	17.0	6.3	11.1
	b	84.5	2.6	10.3	93.2	80.6
	c	11.5	4.2	72.7	**0.6**	8.3
36	a	**0.5**	3.2	9.6	34.7	3.9
	b	5.5	66.7	3.0	61.4	93.9
	c	94.0	30.2	87.3	4.0	2.2
37	a	**2.0**	3.2	23.6	12.5	2.8
	b	92.5	62.6	10.9	84.7	5.6
	c	5.5	34.2	65.5	2.8	91.7
38	a	**2.0**	3.7	75.3	88.6	7.2
	b	86.9	**0.5**	19.3	6.8	85.0
	c	11.1	95.8	5.4	4.5	7.8
39	a	11.6	73.9	77.0	4.0	57.2
	b	78.3	12.2	16.4	**0.6**	13.9
	c	10.1	13.8	6.7	95.5	28.9
40	a	5.0	82.5	**54.5**	13.1	2.2
	b	94.0	6.9	38.8	26.1	5.6
	c	**1.0**	10.6	6.7	60.8	92.2

Table D1 (continued)

Item	Answer choice	Form 1 (*n* = 200	Form 2 (*n* = 189)	Form 3 (*n* = 166)	Form 4 (*n* = 176)	Form 5 (*n* = 180)
41	a	8.0	35.8	4.2	<u>81.1</u>	<u>95.6</u>
	b	4.0	7.5	3.6	14.3	**0.6**
	c	<u>88.0</u>	<u>56.7</u>	<u>92.2</u>	4.6	3.9
42	a	2.5	<u>67.2</u>	**1.2**	3.4	<u>89.4</u>
	b	8.0	12.7	6.0	<u>77.3</u>	**1.7**
	c	<u>89.5</u>	20.1	<u>92.8</u>	19.3	8.9
43	a	4.5	<u>91.0</u>	9.1	2.3	<u>82.2</u>
	b	<u>84.0</u>	**1.6**	**1.8**	<u>94.3</u>	3.9
	c	11.5	7.4	<u>89.1</u>	3.4	13.9
44	a	2.5	14.8	<u>84.9</u>	9.7	2.2
	b	21.1	<u>77.8</u>	**0.0**	<u>85.8</u>	<u>88.3</u>
	c	<u>76.4</u>	7.4	15.1	4.5	9.5
45	a	<u>97.5</u>	6.3	4.8	<u>85.8</u>	3.9
	b	**1.0**	<u>65.1</u>	<u>86.1</u>	9.7	<u>75.6</u>
	c	**1.5**	28.6	9.0	4.5	20.6
46	a	14.0	<u>85.2</u>	9.6	<u>90.9</u>	<u>87.2</u>
	b	<u>78.0</u>	12.7	15.7	**0.6**	8.9
	c	8.0	2.1	<u>74.7</u>	8.5	3.9

Note. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distracter selection rate is too low or too high.

Table D2

Item-Total Correlation for Each Item on Each Form of the English
DL 5T (Rev. 2/01) Test for First-Attempt Provisional Applicants

Item	Form 1 (n = 200)	Form 2 (n = 189)	Form 3 (n = 166)	Form 4 (n = 176)	Form 5 (n = 180)
1	.11	.43	.29	.21	.22
2	.25	.38	.44	.28	.20
3	.14	.36	.44	-.05	.16
4	.26	.13	.22	.04	.36
5	.32	.02	.31	.27	.26
6	.08	.21	.30	.27	.28
7	.19	.25	.45	.21	.27
8	.15	-.08	.30	.24	.25
9	.22	.27	.32	.18	.26
10	.20	.17	.43	.26	.19
11	.31	.26	.32	.22	.08
12	.31	.28	.28	.08	.25
13	.39	.09	.33	.20	.23
14	.39	.13	.07	.02	.44
15	.34	.11	.26	.13	.41
16	.18	.21	.39	.24	.42
17	.36	.21	.31	.30	.18
18	.20	.24	.27	.23	.14
19	.20	.27	.24	.04	.26
20	.22	.33	.34	.26	.14
21	.28	.38	.43	.22	.31
22	.28	.07	.34	.22	.39
23	.24	.34	.22	.33	.28
24	.24	.26	.33	-.03	.26
25	.26	.41	-.05	.14	.19
26	.25	.05	.34	.29	.22
27	.22	.39	.21	.17	.33
28	.33	.20	.42	.29	.39
29	.22	.23	.28	.10	.11
30	.16	.19	.28	.37	.25
31	.10	.12	.18	.13	.28
32	.37	.00	.26	.18	.25
33	.30	.42	.29	.14	.32
34	.23	.24	.39	.28	.12
35	.27	.28	.28	.11	.30
36	.36	.29	.15	-.12	.35
37	.28	.16	.40	.19	.29
38	.33	.30	.38	.30	.30
39	.19	.39	.20	.11	.31
40	.29	.11	.15	.23	.04
41	.32	.03	.34	.10	-.00
42	.25	.36	.47	.18	.31
43	.04	.26	.31	.10	.24
44	.24	.36	.31	.18	.23
45	.37	.17	.29	.35	.20
46	.33	.04	.21	.09	.14

Note. Shading indicates that an item needs to be reviewed and possibly revised or replaced because the item-total correlation is negative or less than .10.

Table D3

Percentage of First-Attempt Original Applicants Who Would Pass if Different
Cut-Points Were Used for Each Form of the English DL 5T (Rev. 2/01)

Number Missed	Form 1 ($n = 200$)	Form 2 ($n = 189$)	Form 3 ($n = 166$)	Form 4 ($n = 176$)	Form 5 ($n = 180$)	Total ($N = 911$)
0	1.0	0.5	0.0	0.0	1.7	0.7
1	3.5	3.7	1.8	0.6	4.4	2.9
2	7.0	8.5	4.8	2.8	7.2	6.1
3	12.5	16.4	7.8	9.7	10.6	11.5
4	21.0	23.3	16.9	14.8	15.0	18.3
5	27.5	34.9	24.1	20.5	23.3	26.2
6	37.5	45.5	29.5	29.5	28.9	34.5
7	48.0	56.1	37.3	36.9	36.7	43.4
8	53.5	67.7	45.2	45.5	45.0	51.7
9	62.5	71.4	55.4	50.0	52.2	58.6
10	72.0	79.4	60.2	56.8	60.0	66.1
11	79.0	82.0	65.1	64.8	67.8	72.1
12	84.5	86.2	71.7	72.2	75.0	78.3
13	87.5	89.4	76.5	80.1	81.1	83.2
14	90.0	92.6	78.9	85.2	82.8	86.2
15	93.5	94.7	83.7	89.2	83.3	89.1
16	94.5	95.2	86.7	92.6	88.3	91.7
17	96.0	96.3	88.6	97.2	90.6	93.9
18	96.5	96.8	90.4	97.7	94.4	95.3
19	98.0	97.4	92.2	98.3	95.0	96.3
20+	100.0	100.0	100.0	100.0	100.0	100.0

Note. The shaded line highlights the pass rates at the current passing standard of eight allowable errors.

Table D4

Summary of Problem Items on Each Form of the English
DL 5T (Rev. 2/01) Test for First-Attempt Provisional Applicants

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items
Item-total correlation too low or negative[a]	6, 43	5, 8, 13, 22, 26, 32, 41, 46	14, 25	3, 4, 12, 14, 19, 24, 36, 46	11, 40, 41
Pass rate too high[b]	7, 10, 11, 15, 24, 45	2, 5, 11, 25, 32, 33, 38	15, 21	15, 19, 39	7, 19, 21, 32, 41
Pass rate too low[c]	4, 12, 29, 31	17, 41	6, 11, 12, 18, 25, 40	9, 30, 33	4, 25, 31, 34, 39
Pass rate too high or low and item-total correlation too low or negative		5, 32, 41	25	19	41
Distracter selected too often[d]	12		25, 40	33	4, 31
Distracter selected too infrequently[e]	1, 7, 10, 11, 12, 13, 14, 15, 24, 30, 32, 33, 36, 37, 38, 40, 45	2, 11, 14, 25, 28, 31, 32, 33, 34, 38, 43,	13, 15, 21, 27, 28, 30, 34, 42, 43, 44	1, 3, 15, 19, 24, 35, 39, 46	7, 11, 16, 19, 21, 23, 28, 32, 41, 42

[a]The item-total correlation was negative or less than .10. [b]More than 95% of applicants answered the item correctly. [c]Less than 60% of applicants answered the item correctly. [d]A distracter was chosen more, or almost, as often as the correct answer to the item. [e]A distracter was selected by 2% or fewer applicants.